

THIS WEEK

EDITORIALS

METRICS ORCID system flowers with added grant information **p.8**

WORLD VIEW Europe looks beyond horizon on gender analysis **p.9**



CHEESE Ancient Chinese tombs give up earliest dairy secrets **p.10**

Careless.data

The authorities must take the necessary time to remedy the slapdash introduction of a database containing the medical records of the entire population of England.

England's universal health-care system has potentially an enviable ability to integrate medical records into a data system that benefits patients and researchers alike. What a shame that it is making such a hash of the task, and undermining crucial public and professional trust in the process.

The National Health Service (NHS) in England has postponed the launch of its controversial care.data programme for six months, abandoning its earlier plan to start uploading data from this spring. The sensitive data consist of previously confidential records from people's consultations with their doctors, such as their family history, diagnoses, prescriptions and the results of blood or other tests.

The data store, managed by the NHS's Health and Social Care Information Centre (HSCIC) in Leeds, is part of an ambitious scheme to link patients' doctors' records with their hospital and other medical data, to create one of the world's most comprehensive health-care databases. The plan is initially for the NHS to use the data to improve health-care management, with researchers and other users eventually being allowed access.

But the programme has been controversial, raising serious concerns, highlighted by this journal in January (see *Nature* **505**, 261; 2014), that the authorities have paid lip service to informed consent, leaving the public in the dark about how their personal data would be safeguarded and used, and obfuscating their right to opt out. Opposition from patients, doctors and campaign groups swelled into open revolt in early February, with key bodies, such as the British Medical Association and the Royal College of General Practitioners, joining the chorus of criticism that the programme was being rushed.

On 19 February, the NHS, which had stubbornly dismissed such criticisms, reluctantly admitted that its critics might have a point, and postponed the launch until the autumn. But the agency still seems to be in denial, arguing that opposition is merely down to problems of communication and public perception. That response is wrong. Care.data is far from ready for launch.

PUBLIC TRUST

Incredibly, with just a few weeks to go to the scheduled launch, the NHS had not even laid out in detail which groups would be able to access the data, and on what terms. Thus, the public could not fully know what they were signing up to, raising fears that personal data might end up, for example, in the hands of insurance and other commercial companies. Economic growth is a core goal of any UK government, but there has to be confidence that this won't trample on the rights of individuals.

Adding to the mess, over the past week it has emerged that the Leeds centre has handed over records from its existing massive database of personal hospital data to outside groups in ways that might have contravened its own rules. At the very least, these episodes raise questions about the functioning of the HSCIC, but they also risk diminishing what little public confidence is left in care.data.

Moreover, at a time when there has never been a greater need for the operators of care.data to win public trust, the HSCIC has stonewalled about explaining the terms and conditions of the data transfers made. It must quickly disclose full details of the transfers, and of any past violations.

Jeremy Hunt, the UK health secretary, seems to have grasped that the row over care.data is not just a public-relations hitch, and that the haphazard planning puts the scheme at risk of going into meltdown, with the loss of many of its benefits. Last Friday, he announced his intention

"Care.data needs to be thought about more deeply and recrafted, fully engaging all stakeholders."

to legislate to prevent the HSCIC from releasing pseudonymized data — in which patients' NHS numbers are replaced by pseudonyms, but where people can often be re-identified — unless there were clear health benefits, and to ban the release of data for commercial purposes. He also said that the HSCIC would be subject to new external statutory oversight.

Hunt's move is a step in the right direction, but the programme needs to be thought about more deeply and recrafted, with the necessary time given to engage all stakeholders more fully, including researchers and opponents of the scheme.

Whatever laws are introduced in England, they will be subject to an upcoming European Union (EU) General Data Protection Regulation, which will be legally binding on member states and so override national laws. The EU law, a major revision of 1995 legislation, also governs the use of medical records, and last October, the European Parliament's Committee on Civil Liberties, Justice and Home Affairs voted for amendments that would remove many of the exemptions for medical research.

Although anonymized data could be used without restriction, one amendment would allow the use of pseudonymized data without consent only if the research was of "high public interest" and could not be carried out otherwise, with requests needing to be reviewed by an independent body. A joint statement in January by dozens of European research organizations and charities warned that medical research would be "severely threatened" by this, and pointed out that the use of such data is already subject to strong institutional ethical oversight.

Although the use of personal data for research, with the attendant high ethical standards, is probably the least of the worries, scientists must be sensitive to EU public concerns about the collection and use of health data more broadly. As always, democratic scrutiny, and striking a balance between regulatory overheads and genuine concerns, will be key. Again, it will be important to take the time to have the debate needed to get this balance right.

It is a fundamental human right that people can determine how their personal medical data are used, and exceptions to specific informed consent cannot be taken for granted. Informed consent is not an obstacle to be overcome, but a principle to be respected and cherished. ■

Full disclosure

Regulatory agencies must demand conflict-of-interest statements for the research they use.

It was the 1976 film *All the President's Men*, about the uncovering of the Watergate political scandal by two *Washington Post* reporters, that popularized the phrase: "Follow the money." He who pays the piper calls the tune. Science combats the undue influence of commercial interests — or at least tries to — by using a different guideline, illustrated by a popular catchphrase from another film: "Show me the money." Give us transparency.

The selective promotion of scientific research to steer policy-making is a murkier business altogether — particularly in environmental policy-making, in which the battle for the ear of the piper between big business and the 'little guy', who is often affected by pollution or hazardous substances, is so asymmetric. The problem is not limited to climate change, which is only the most high-profile example at present.

It has been more than a decade, for example, since David Michaels, previously a public-health researcher at George Washington University in Washington DC, and Wendy Wagner, an environmental-law specialist, broached the issue in the pages of *Science* (D. Michaels and W. Wagner *Science* **302**, 2073; 2003). They warned that the evidence base of important regulatory standards is undermined by the limited scrutiny of private research submitted to regulatory bodies, and by the fact that these bodies often do not require disclosure of researchers' funding sources.

Michaels is now in a position to do something about this. In 2009, he was appointed to lead the Occupational Safety and Health Administration (OSHA), one of the US agencies he criticized in that 2003 piece.

OSHA's remit is health-and-safety standards, and the test bed for Michaels' stance is a 40-year effort to regulate exposure to silica dust. Crystalline silica dust is produced by processes such as concrete grinding and sandblasting in construction and other industries. If inhaled, it can cause silicosis — an incurable condition

involving inflammation of the lungs — and lung cancer.

As part of a consultation on tougher regulation of silica exposure, OSHA asked that people submitting scientific comments to the agency should declare financial conflicts of interest. According to Michaels, this might be the first time that any federal agency has made such a request.

But even though this is a request and not a requirement, it has not gone down well in all quarters. In particular, a group of powerful US senators has come out against the idea that such a declaration should be part of federal rule-making (see page 18).

"There is a broad consensus in favour of transparency about funding sources."

They suggest that OSHA might "prejudge the substance" of comments on the basis of such disclosures.

Nature — like many journals — has required such disclosures for years, and considers such opposition to be misguided. In controversial areas, these conflict statements pre-empt allegations of secrecy and bias that could distract from the central issues. And past failure to be transparent about such interests has led to scandals involving concealed or distorted evidence and ghost-writing, as has been well documented in areas from tobacco control to drug development.

The medical profession and the pharmaceutical industry, to their credit, have taken major steps towards openness. Some researchers think that conflict-of-interest disclosures should go even further than they currently do, and should detail the contractual arrangements involved, such as whether the funder had a veto on publication. In science more generally, there is a broad consensus in favour of transparency about funding sources.

Transparency is the best defence against the purchase of undue influence by those with the most financial clout. In areas where tough standards are needed to protect public health, and powerful and wealthy interests have a financial incentive to water down these standards, such transparency is more than desirable — it is essential, and history demonstrates that. Rather than challenging OSHA for requesting conflict-of-interest disclosures, US politicians should be asking why all federal agencies do not require them. After all, it is easier to follow the money, and to make the proper decision, when all details are on full show. ■

Track and trace

Identifiers that follow researchers' work from grant to paper will make funding more effective.

More than half a million researchers have now signed up for an online science passport: a unique 16-digit identity number, with an accompanying online profile, from the Open Researcher and Contributor ID (ORCID) project. There, researchers can maintain an up-to-date record of their professional pursuits.

Already, ORCID is being integrated into the ecosystem of science: many publishers accept ORCID identifiers in their manuscript-submission processes, and funders including the Wellcome Trust and the US National Institutes of Health are accepting the identifiers to streamline grant applications. Universities and research institutions are planning to use the system to track their researchers' output throughout their careers.

So far, the ORCID website has prompted scientists to record outputs such as articles, data sets, citations, patents and media appearances. This fits in with the growing desire of institutions and funding agencies to recognize the full range of researchers' activities and impacts.

But this week, ORCID begins to request a new set of data — inputs. Researchers logging in to their profiles will be prompted to add the

details of their grants, or to confirm information on grants they hold. Such information is often publicly available on the Internet, but scattered across funding-agency websites, rather than collated for individual scientists. ORCID hopes to improve tracking of the connections between the cash that funders pour into research and the results that emerge.

Another service that makes it easier to link grants in with papers out is FundRef, launched last year by the non-profit publisher alliance CrossRef. It provides a standardized format for adding funding information to the metadata of research articles published online.

The result — if such systems catch on — should be easier tracking of the efficiency of the science system. Which academics produce the most for the grants they receive, and why? What kinds of grants are most effective at prompting what types of output? That is something funders and economists would dearly like to know. They have made individual efforts, but a bigger-picture understanding has been held back by lack of connectivity across agencies.

There is perhaps a danger that scientists — so used to measuring the properties of others — will be resistant to having information recorded on themselves. (Less than one-quarter of researchers signed up to ORCID have actually listed at least one output on their profiles.) But ORCID (of which Nature Publishing Group is a partner) gives researchers control over the information that they allow to be publicly visible. Hopefully, they will embrace the opportunity to make science funding more effective and evidence-based. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv



Scientific research must take gender into account

From car design to drug discovery, the failure to acknowledge sex differences can be costly and even lethal, argues Linda Schiebinger.

In Madrid a couple of years ago, I was interviewed for Spanish newspapers. When I later ran the text through Google Translate, I got a shock: I was referred to repeatedly as “he”.

Like much science and technology, Google Translate has a male default. When I drive a car, the seatbelt is not designed to accommodate breast tissue. Any medicines I take are more likely to have been tested on male than on female animals. There are moral issues here: women pay taxes and buy products and should not be short-changed. But scientific objectivity is at stake, too. Because medical research is done mainly in males, there is a male bias in, for example, the choice of drug targets. Science is halving the potential field of innovation.

This is not about active discrimination; the bias is largely unconscious. Google Translate defaults to the masculine pronoun because ‘he’ is more commonly found on the Web than ‘she’. Yet that is changing: an analysis of American-English texts in Google Books shows that the ratio of masculine to feminine pronouns has fallen to around 2:1, from a peak of 4:1 in the 1960s.

In the summer of 2012, I invited Google and several language-processing experts to a Gendered Innovations workshop at Harvard University in Cambridge, Massachusetts. They listened to the problem for about 20 minutes, then said: “We can fix that!” Although it is complicated, the search for solutions is on. Fixing the problem is great, but constantly retrofitting for women is not the best road forwards.

A better way is to include gender at all relevant phases of research — when setting priorities, gathering and analysing data, evaluating results, developing patents and, finally, transferring ideas to markets. Science and technology should take into account the biological and social needs of both women and men.

Unconscious sex and gender bias can be socially harmful and expensive. In automotive engineering, short people (many women, but also many men) are classed as ‘out-of-position’ drivers and often ignored. This leads to greater injury in accidents. In medicine, osteoporosis has long been defined as a disease mainly of post-menopausal women — an assumption that has shaped screening, diagnosis and treatment. Yet after the age of 75, men account for nearly one-third of osteoporosis-related hip fractures. And in basic biomedical research, the failure to use female cells, tissues and animals can lead to greater health risks for women. Of the ten drugs withdrawn from the US market between 1997 and 2000, eight posed greater threats to women than to men. Developing a drug costs billions of dollars, and failure can cause human suffering and death — with stakes this high, why ignore half of the population?

In December last year, the European Commission took a historic step towards ending this

unconscious bias. Applicants to its newly opened Horizon 2020 funding scheme are now asked to include gender analysis in their projects — for example, to assess whether the research will have different implications for women and men. The commission identified dozens of science areas that could benefit from gender analysis: computer hardware and architecture, biodiversity, ecology, biophysics, oceanography, geosciences, organic chemistry, aeronautics, space medicine and some 40 others, including nanotechnology (astrophysics did not make the cut).

Since 2006, the Canadian Institutes of Health Research has required scientists across its 13 institutes to analyse sex and gender (when appropriate); and since 2008, the philanthropic Bill & Melinda Gates Foundation has required gender analysis for its agricultural research grants. The European Commission’s move with Horizon 2020 is the most significant in terms of scope.

Resistance sometimes comes from those who ask: won’t including females in animal studies increase costs? For an individual laboratory, it probably will. But removing bias from science will cost society less in the long run — and save lives. How can we safely include women in phase III human trials (as required by US law since 1993) if drugs are not first tested in female laboratory animals?

Many of these problems have been discussed before. But there are signs that more people are taking the issue of gender analysis seriously. Publishers, for example, are catching on and insisting that published work accounts for gender. *Clinical Orthopaedics and Related Research* recommends

that studies be sufficiently powered to analyse sex and gender, and in 2012, each of the American Physiological Society’s 14 journals required that authors report and analyse sex effects. More journals should follow.

Including gender analysis in research can save us from life-threatening errors... and can lead to new discoveries. Gender analysis has led to better treatments for heart disease in women. Identifying the genetic mechanisms of ovarian determination has enhanced knowledge about testis development. Analysing how sex affects donor-recipient matching is improving stem-cell therapies. And exploring how sex-specific biological factors and gender-specific behaviours interact has helped researchers to understand how nutrients trigger cell functions, and may assist in the fight against obesity.

Can we afford to ignore such opportunities? ■

Linda Schiebinger is the John L. Hinds professor of history of science at Stanford University, California, USA.
e-mail: schieb@stanford.edu

INCLUDING
GENDER
ANALYSIS IN
RESEARCH CAN SAVE
US FROM
**LIFE-
THREATENING**
ERRORS.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/n6pglw

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CANCER

How cancer skirts brain defences

Proteins that block cell death and help cells to integrate with blood vessels are crucial in the spread of cancer to the brain.

Brain metastasis is often deadly, but most cancer cells that invade the brain die without establishing a tumour. To find ways in which successful invaders bypass the brain's defences, Joan Massagué of the Memorial Sloan Kettering Cancer Center in New York and his colleagues looked at a set of genes expressed in brain metastases, focusing on two proteins called serpins.

These serpins inhibit another protein called plasminogen activator, which, the authors found, kills cancer cells in the brain. Serpin expression shielded cancer cells from cell death and helped them to spread on the surface of capillaries, establishing a blood supply in their new home.

Cell 156, 1002–1016 (2014)

MATERIALS

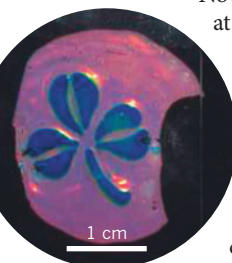
Changing colour under pressure

Materials that change colour when pulled or squeezed could form the basis of display screens or sensors. But existing photonic gels, which change colour when deformed, cannot cover the entire rainbow or switch quickly.

Now, Jianping Ge at East China

Normal

University in Shanghai and his colleagues have created a material that can adopt every colour from red



ARCHAEOLOGY

Ancient cheese found with mummies

The oldest known pieces of cheese have turned up in the tombs of an early Bronze Age cemetery in Xinjiang, China.

Andrej Shevchenko at the Max Planck Institute of Molecular and Cell Biology and Genetics in Dresden, Germany, Changsui Wang at the University of Chinese Academy of Sciences in Beijing, and their colleagues analysed 3,800-year-old lumps found at the neck and chest of mummies (pictured) in the cemetery and identified them as a 'kefir' cheese.

This type of cheese is made by curdling

ruminant milk with a symbiotic culture of bacteria, including *Lactobacillus kefiranofaciens*, and yeast. Evidence of a kefir dairy — which makes lactose-free products — in this region explains why large-scale ruminant herding and milking spread in a population known to have been lactose intolerant, the authors say. The origin of cheese making dates back some 4,000 years earlier, but evidence for this has relied on analysis of milk fat in pottery shards.

J. Arch. Sci. <http://doi.org/rpq> (2014)

GENETIC ENGINEERING

Genes make bacteria magnetic

Researchers have transferred genes for the production of magnetic nanocrystals from one species of bacteria to another, a step towards making bacterial bioreactors that generate such particles.

Dirk Schueler at Ludwig-Maximilians University in Munich, Germany, Youming Zhang at the

Helmholtz Joint Institute at Shandong University in Jinan, China, and their team focused on roughly 30 genes from *Magnetospirillum gryphiswaldense* that enable the bacterium to produce membrane-bound, iron-based magnetic nanocrystals. Researchers inserted these genes into *Rhodospirillum rubrum*, a well-studied organism used in biotechnology, that is easier to work with. The authors found that both sets of magnetic

YUSHENG LIU

DONGPENG YANG

particles were similar in size, structure and composition.
Nature Nanotech. <http://doi.org/rpq> (2014)

MICROBIOLOGY

Stubborn microbe finds hiding spots

Salmonella bacteria can escape antibiotics and immune-system attack by hiding inside a host's immune cells.

Roland Regoes and Wolf-Dietrich Hardt of the Swiss Federal Institute of Technology in Zurich and their colleagues infected mice with *Salmonella enterica* and then treated the animals with the antibiotic ciprofloxacin. The team found that, after most of the infection had cleared from organs, about 10% of the *Salmonella* bacteria inside lymph nodes that drain the intestines were still viable and growing. These bacteria re-established infection after antibiotic treatment.

In a separate study, Sophie Helaine, David Holden, and their colleagues at Imperial College London found that *Salmonella* cells can also persist inside macrophages that ingest them. Molecules that stimulate certain types of immune cells could, in combination with antibiotics, improve treatment.
PLoS Biol. 12, e1001793 (2014);
Science 343, 204–208 (2014)

ANIMAL BEHAVIOUR

Worms bond to reach new heights

To reach passing beetles, parasitic worms congregate into towers up to 30 times taller than an individual.

Hans-Joachim Knölker at the Dresden University of Technology in Germany, Teymuraz Kurzchalia at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, and their team found that juvenile *Pristionchus pacificus* secrete a waxy substance, dubbed nematol by the authors, that functions as an adhesive

between animals. This allows them to form towers with up to 1,000 individuals, reaching as high as 1 centimetre, which can attach to a host.

Nature Chem. Bio. <http://doi.org/rqt> (2014)

MATERIALS

Waste glass finds diffuse use

A by-product of antiquated glass-making processes could find a new lease of life as an optical diffuser.

Manufacturers now avoid producing crystals, called devitrite, in their soda–lime–silica glass because they degrade the material's optical qualities. But Haider Butt at the University of Birmingham, UK, and his team show that this currently undesirable form of glass can act as a diffuser, owing to the fact that needle-like crystals of devitrite scatter light to wide angles of up to 120°. Because devitrite is both a highly efficient diffuser and cheap to produce, the authors say it could be used in applications such as medical lasers.

ACS Nano. <http://doi.org/rpk> (2014)

PALAEOLOGY

Algae dealt blow to ancient whales

The unearthing of more than 40 marine mammal fossils (pictured) at a site in Atacama, Chile, has revealed that they probably died en masse in four events due to toxic algae.

Nicholas Pyenson of the Smithsonian Institution in

COMMUNITY CHOICE

The most viewed papers in science

NEUROSCIENCE

Hopping DNA linked to schizophrenia

HIGHLY READ
on www.cell.com
27 Jan–26 Feb

Mobile DNA elements activated by environmental or genetic triggers could boost susceptibility to schizophrenia.

Tadafumi Kato at the RIKEN Brain Science Institute in Saitama, Japan, Kazuya Iwamoto at the University of Tokyo and their colleagues found higher copy numbers of a genetic element, called the L1 retrotransposon, in the DNA extracted from the post-mortem brains of patients with schizophrenia compared with that from the brains of healthy individuals. The authors also showed that the L1 elements tended to localize to genes linked to schizophrenia and neuronal synapses.

The same phenomenon was observed in mice and monkeys exposed perinatally to chemical stressors that are known to promote schizophrenia-like behaviours, and in cultured neurons derived from stem cells carrying a schizophrenia-related chromosomal deletion.

Neuron 81, 306–313 (2014)

Washington DC and his colleagues found that the fossils, including baleen and other whales, seals and an aquatic sloth, were arranged in four distinct layers, which are between 9 million and 6.5 million years old. The only modern event known to trigger such recurring and rapid die-offs of multiple species is extreme growth of toxic algae. The animals could have died after eating contaminated prey or have been directly affected by the algae's toxin. Other ocean areas that once created such blooms could harbour substantial fossil vertebrate remains, the authors say.
Proc. R. Soc. B 281, 20133316 (2014)

ZOOLOGY

Pheromone turns on goat brains

Researchers have pinpointed a molecule produced by male goats that activates reproduction in females outside of their normal breeding season.

Yukari Takeuchi at the University of Tokyo and her colleagues used a special cap to capture 18 different pheromone molecules emitted from the heads of male goats. They then exposed females to the molecules, and used implanted electrodes to monitor activity in a brain region that regulates reproduction. One molecule, 4-ethyloctanal, triggered the biggest response of all the compounds tested.

This is the first molecule shown to stimulate a key regulator of reproduction in mammals called the gonadotropin releasing hormone pulse generator.
Curr. Biol. <http://doi.org/rqv> (2014)

NATURE.COM

For the latest research published by Nature visit:
www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

Research relegation

Switzerland-based scientists and companies will have restricted access to European research funds in 2014. Switzerland, which is not a member of the European Union (EU), has now lost its status as an associate partner in the EU's €80-billion (US\$110-billion) Horizon 2020 funding programme, after it imposed curbs on immigration last month (see *Nature* 506, 277; 2014). The relegation means that research institutes and companies based in Switzerland are no longer automatically eligible to apply for EU funding. Moreover, individual scientists cannot use basic-research grants from the European Research Council at Swiss institutes.

US Navy looks north

By 2030, the Northern Sea Route — a shipping lane running through the Arctic — could be ice-free and navigable for at least nine weeks each year, the US Navy said on 24 February in its 'Arctic roadmap'. The Northwest Passage — an alternative shipping route through the Arctic — could be open five weeks each year, the roadmap says. It also projects that the region's waterways will see rising activity from fishing, tourism, and oil and gas exploration, but says the development poses little risk to US national security. See go.nature.com/fv9skt for more.

DNA swap

The United Kingdom could become the first country to legalize mitochondrial replacement — a reproductive technology that produces offspring with three genetic parents. The technique could prevent children from inheriting diseases that

affect mitochondria, the cell's energy producers, by transplanting nuclear genetic material to a donor cell with healthy mitochondria. A public consultation on the draft legislation is open until 21 May. In the United States, a Food and Drug Administration advisory panel examined the science behind the technologies on 25–26 February. See go.nature.com/ho8haz for more.

ITER critique

The governing council of ITER, a multibillion-euro international nuclear-fusion experiment, has approved 11 urgent reforms to the project's management. The executive summary of a critical assessment report considered at the council's 13 February

meeting was leaked online last week by *The New Yorker* magazine (see go.nature.com/7ayyqe). It details long-standing problems with ITER's complex international management (see *Nature* <http://doi.org/nwqj>; 2012), but is more scathing than past assessments, calling for an "achievable and realistic schedule". It adds that the "schedule has slipped so consistently ... almost no one ... believes the announced dates for first plasma [2027] or other key milestones will be met".

Leaky helmet

An incident that nearly drowned an astronaut on a spacewalk last July could have been avoided, says a NASA report published on 26 February. Italian astronaut

The satellite was built by NASA and the Japan Aerospace Exploration Agency (JAXA), and the data it will collect could aid the accuracy of forecasts of severe weather, estimates of freshwater supplies and projections of climate change. See go.nature.com/rf9k5s for more.



NASA/JAXA

Snow satellite launch

A joint US–Japanese mission to monitor rain and snow launched from the Tanegashima Space Center in Japan on 27 February. The US\$933-million Global Precipitation Measurement Core Observatory (pictured) will map precipitation across 90% of the globe every three hours.

Luca Parmitano had to feel his way back to the International Space Station airlock as more than 1 litre of water collected in his helmet, covering his eyes and nose. The water came in through a leak in his suit. A week earlier, the same leak had occurred, but the crew mistakenly thought that the water came from his drinking tube. The 'close-call' spacewalk would have been postponed if astronauts aboard the space station had fully investigated the first incident.

Radiation leak

The US Department of Energy reported on 26 February that 13 employees had tested positive for low-level radiation exposure following a leak at its Waste Isolation Pilot Plant near Carlsbad, New

Mexico. The employees were above ground at the time of the 14 February leak, but preliminary tests suggest that they inhaled radioactive particles. Energy department officials said the health and environmental impacts of the leak seem to be minor, but traces of plutonium and americium were detected as far as 1 kilometre from the site. See page 15 for more on US nuclear waste storage.

PEOPLE

Final verdict

The South Korean Supreme Court upheld on 27 February a 2010 ruling that sentenced disgraced cloning researcher Woo Suk Hwang to an 18-month prison term for embezzlement and violation of the country's bioethics law. But Hwang will serve jail time only if he breaks laws during a two-year probation period. The court also annulled an earlier decision that would have forced Seoul National University to reinstate Hwang. This is the final judgment in a trial that started in 2006. See go.nature.com/tp7tq9 for more.

Borlaug bronze

A bronze statue of Norman Borlaug (pictured), the plant scientist who won the Nobel Peace Prize in 1970 for developing high-yielding



wheat crops, will be unveiled at the US Capitol building in Washington DC on 25 March. The state of Iowa — where Borlaug was born — will install the statue, which depicts him taking field notes, to mark the 100th anniversary of his birth. Each US state is allowed to place two statues of prominent citizens in the National Statuary Hall Collection.

PUBLISHING

Gibberish papers

The publisher Springer will remove 16 computer-generated nonsense papers that it had published in its subscription database, it said last week. The papers, created by a computer program called SCIdgen, were spotted by computer scientist Cyril Labbé of Joseph Fourier University in Grenoble, France. Last week, *Nature* revealed that Labbé had privately informed Springer of the problem. He had also

contacted the US Institute of Electrical and Electronic Engineers (IEEE), which has already withdrawn more than 100 nonsense articles. The papers had been included in conference proceedings, some of which were supposed to be peer-reviewed. See go.nature.com/d7h4sr for more.

Data sharing

The Public Library of Science (PLOS), an open-access publisher based in San Francisco, California, has strengthened its policy requiring authors to make data publicly available. From 3 March, researchers publishing in PLOS journals must share the data underlying their findings as soon as their manuscript is published and must do so on the Internet, for example in public data repositories. Some exceptions are allowed, for instance in cases in which patient privacy would be compromised.

RESEARCH

Ageing data

The US National Institutes of Health (NIH) announced on 26 February that it had added genetic and medical data from more than 78,000 people with an average age of 63 to an online genetics database, as part of its Genetic Epidemiology Research on Aging project. The data,

COMING UP

10–11 MARCH

The Royal Society in London hosts a meeting on phylogeny, extinction risks and conservation to discuss how information on evolutionary relationships can be combined with metrics such as lists of endangered species to assess conservation needs.

go.nature.com/pwwz8i

11 MARCH

The US Department of Agriculture ends its extended public comment period on whether it should allow sales of maize (corn) and soya bean seeds that are genetically engineered to resist herbicides such as 2,4-D.

collected with US\$25 million of NIH funding, will enable genome-wide association studies that will search for genetic signatures linked to a variety of diseases, including those associated with ageing.

Repatriated bones

Two German universities this week handed over 32 human skulls and three skeletons of Namibian origin to Namibian officials. The bones derive from tribe members who were killed by German troops or who died in concentration camps in the early twentieth century when Namibia was a German colony. Scientists used them for now-discredited research to determine anatomical differences between races. Researchers identified the remains among thousands of bones in historic anthropological collections at the University of Freiburg and the Charité Medical University in Berlin.

► NATURE.COM

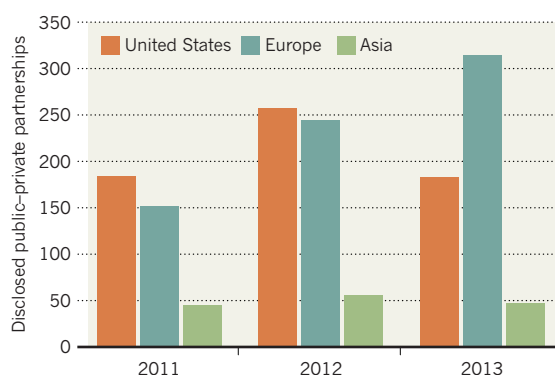
For daily news updates see:
www.nature.com/news

TREND WATCH

Biotech firms are increasingly seeking partnerships with European academic institutes and funders, with 314 publicly disclosed links in 2013. The European Commission and Cancer Research UK made the most links, according to an analysis by the publication *SciBX*. Among companies, AstraZeneca had the most public-private partnerships, including its deal with Sweden's Karolinska Institute to create a US\$100-million cardiac-research centre. See go.nature.com/2fwpsq for more.

PUBLIC-PRIVATE PARTNERSHIPS SOAR IN EUROPE

Links between drug companies and funders or institutions are being forged in Europe more than in the United States.



Data include double counting because some partnerships involve companies and/or institutions from more than one region.

NEWS IN FOCUS

ENERGY POLICY Three years after Fukushima, Japan is at nuclear crossroads **p.16**

GENETICS New prenatal screens may mean hard choices for expectant parents **p.19**

PALAEONTOLOGY Brazil cracks down on illegal fossil trade, rekindling export debates **p.20**

TECHNOLOGY Researchers strive to kick the lithium-ion battery from its throne **p.26**



The Waste Isolation Pilot Plant's salt beds hold radioactive materials from US nuclear-weapons labs.

NUCLEAR POLICY

US seeks waste-research revival

Radioactive leak brings nuclear repositories into the spotlight.

BY JEFF TOLLEFSON

A radiation leak has raised questions about the safety of the United States' only deep nuclear-waste repository, and has given fresh voice to scientists calling for more research into underground waste storage.

On 14 February, radioactive plutonium and americium leaked out of the Waste Isolation Pilot Plant (WIPP) near Carlsbad, New Mexico, where thousands of drums of contaminated material from the US nuclear-weapons programme are stored in salt beds more

than half a kilometre below the surface. The health and environmental impacts seem to be minor, but 13 employees have tested positive for low-level contamination. The Department of Energy (DOE) and its contractors are still working on a plan to re-enter the WIPP and find out what caused the leak.

The incident also brings renewed attention to a problem that policy-makers have been avoiding: what to do with a mounting stockpile of spent fuel from commercial reactors, which is currently stored at reactor sites. In 2010, the DOE mothballed plans to develop Yucca Mountain in Nevada, which

since 1987 had been designated as the future site of an underground repository (see *Nature* **473**, 266–267; 2011). Researchers at the DOE and universities want to explore a variety of alternatives. But they say that they have been hobbled by small budgets and the Nuclear Waste Policy Act, which prevents the DOE from investigating any specific site apart from Yucca Mountain.

“Basically, all of the old ideas have come back out of the woodwork,” says Michael Driscoll, a nuclear engineer at the Massachusetts Institute of Technology in Cambridge. “But the first thing we need is Congress to wrestle with this and revise the Nuclear Waste Policy Act.”

For now, researchers are pursuing generic repository science that does not conflict with the law. In one large proposed experiment, DOE scientists wanted to assess whether salt beds at the WIPP could store radioactive waste that is hotter than the material they currently hold. In 2011, the team began developing a US\$31-million experiment that would have tested how the salt deforms when it is heated, and how water moves through it.

Other researchers are investigating the concept of dropping cylinders of nuclear waste into 5-kilometre-deep boreholes in hard rock such as granite. Sandia National Laboratories in Albuquerque, New Mexico is leading a consortium of researchers and companies seeking to drill an experimental borehole costing approximately \$25 million. The hot-salt and borehole proposals are now competing for funding within the DOE's relatively small \$15-million annual budget for this kind of research. “Big tests like either of those would completely overwhelm the current budget,” says Peter Swift, who heads the DOE's nuclear-waste science programme at Sandia.

In Europe, scientists have developed expertise with other types of rock. Finland and France have homed in on proposed underground repositories in granite and shale, respectively. Germany has buried low- and medium-level wastes in underground domes of salt, and it is evaluating the terrain for a controversial high-level waste repository.

International collaboration gives researchers access to the basic science on all of these environments, says Jacques Delay, secretary-general of the Implementing Geological Disposal of Radioactive Waste Technology ►

► Platform in Bure, France, a consortium that guides a roughly €10-million (US\$14-million) joint research programme under the European Commission. “What is tricky is to make the link between the academic science and our projects,” he says.

But basic research can go only so far, because the scientific assessment of repository safety is specific to local geology. After choosing a site, researchers must study the density, porosity and heat conductance of the rock there, and characterize any fractures and groundwater movement. Modelling and experiments help to determine how the rock will respond to the heat generated by the nuclear waste.

The United States spent more than \$15 billion on Yucca Mountain before then-energy secretary Steven Chu pulled the plug, saying that the site was not a “workable option” — broadly interpreted to mean that it was doomed politically, if not technically. The United States has evaluated few alternatives. The city of Carlsbad, which hosts the WIPP, is politically inclined to expand its nuclear-waste portfolio. But few other communities have shown interest in storing nuclear waste.

Some DOE researchers favour a serious exploration of borehole disposal, in part because no one has tested the idea, which dates back to the 1970s. Swift estimates that just 800 boreholes would take care of the existing US waste stockpile, as well as spent fuel from current reactors until about 2050. There is suitable rock at various depths across the country. “You could spread these things out, and you wouldn’t have to put all of your money on one site,” says Patrick Brady, a geochemist at Sandia who is part of the lab’s borehole consortium.

Drilling constraints might limit these boreholes to less than 50 centimetres in diameter, so spent fuel rods, currently stored in large canisters, would need to be repackaged. However, a hole that size would be perfect for a major source of waste that the DOE is trying to dispose of: 2,000 highly radioactive capsules containing caesium and strontium from the Hanford Site, a decommissioned plutonium-production facility in Washington state. These capsules are 52–56 centimetres long and up to 9 centimetres in diameter, and they contain 38% of Hanford’s radioactivity. Swift says that they could all fit into a single borehole.

With research worldwide concentrating on underground repositories, Swift says that it is time to try a new concept: “If we make a borehole, it will be the one that the rest of the world comes and looks at.” ■



An abandoned train line in Namie, Japan, inside the restricted area around the Fukushima Daiichi plant.

TOMAS MUNI/ATNY TIMES/REDUX/EYEVINE

ENERGY POLICY

Japan caught up in energy dilemma

As the third anniversary of the Fukushima disaster nears, the nation is faltering in its resolution to shun nuclear power.

BY DAVID CYRANOSKI

Three years after a tsunami led to reactor meltdowns at Japan’s Fukushima Daiichi nuclear power station, the country is at a crossroads in terms of energy policy. A draft law released last week suggests that, despite the previous government’s promise of a ‘zero-nuclear’ future in the wake of the disaster, the current administration is ready to re-embrace the technology. Yet the announcement came just weeks before the opening of a ¥10-billion (US\$98-million) renewable-energy research centre in Fukushima prefecture that aims to be at the forefront of green technology. Which way will Japan turn?

The reformist wave that swept Japan after the 2011 disaster included proposals to supply all of the country’s energy from renewable sources. Nowhere is taking this more seriously than Fukushima prefecture, which plans to use an array of giant solar panels, biomass plants and windfarms to supply all of its energy by 2040. Two floating, 7-megawatt wind turbines, the world’s largest, are scheduled to come online in the next year.

The opening of the Fukushima Renewable Energy Institute in April will bolster the prefecture’s vision. The institute is in Koriyama, 60 kilometres west of the stricken nuclear plant, and has been established by the Tsukuba-based National Institute of Advanced Industrial Science and Technology (AIST). It has attracted interest from electronics companies such as Panasonic and Sharp, and foreign collaborations

are also in the pipeline, including one with the Fraunhofer Institute for Solar Energy Systems in Freiburg, Germany. Inside, about 100 researchers will work across areas including solar, hydrogen, wind and geothermal power.

Deputy director Tetsuo Munakata says that the institute will expand on the long-running research programmes of AIST, and he points to concrete goals such as reducing the thickness of silicon wafers for solar cells to 80 micrometres in seven years.

But some see the institute as a showpiece with little chance of success. Critics point out that Japanese solar-panel makers are already struggling to compete with the lower costs of Chinese manufacturers. Tetsunari Iida, head of the Institute for Sustainable Energy Policies in Tokyo, says that the Fukushima institute lacks experienced hands and will get bogged down with ministerial bureaucracy. “I don’t think we can expect much in terms of practical results,” he says.

Sceptics also point to what seems to be a government move towards nuclear energy. Despite continued leaks of radioactive water stored at Fukushima Daiichi, and extensive exclusion zones remaining around it (see ‘Fukushima: the legacy’), the draft energy plan says that the government will push to restart Japan’s 48 operable reactors, all of which were closed after the earthquake that triggered the disaster. The plan is expected to be approved in the spring.

But Munakata says that the new institute also underlines the government’s commitment to renewable energy. “There’s no way it isn’t going to keep investing in renewables,” he says. ■

FUKUSHIMA: THE LEGACY

The region around the Fukushima Daiichi nuclear power plant is still subject to exclusion zones that roughly mirror the areas of greatest contamination. This map shows the perimeters and radioactivity 1 metre above the ground in September 2013.

CANCER RISKS
Radiation exposure in Fukushima poses only a negligible cancer risk. For example, a study last month showed that residents in the worst-affected residential areas, including Tamano, have only a 1% higher lifetime risk of developing solid cancers.



OFFSHORE RADIATION
In December 2013, the International Atomic Energy Agency found that the level of caesium-137 in the ocean around the Fukushima plant was 1 becquerel per litre (Bq L⁻¹). World Health Organization guidelines state that the maximum level of caesium-137 in drinking water should be 10 Bq L⁻¹.

OCEAN POLLUTION
The radiation levels farther offshore, between 2 and 20 kilometres from the port, are now mostly below 0.1 Bq L⁻¹ for caesium-137. Even farther out, the levels are close to what they were before the disaster: 0.001–0.003 Bq L⁻¹ for caesium-137.

RADIATION DOSES
According to the World Health Organization, residents living in the hardest-hit areas of Fukushima prefecture were subject to radiation levels of 12 to 25 millisieverts (mSv) in the first year. That is equivalent to two or three medical CAT scans.

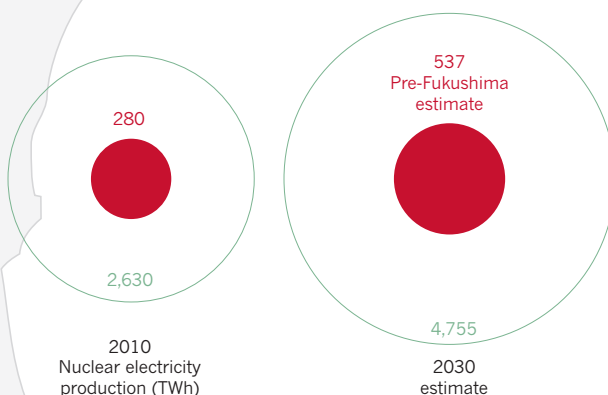
EXCLUSION ZONES
Officially, 154,000 people have been evacuated from their homes near the plant. None have yet been allowed to return permanently, but over the next two years up to 30,000 may be allowed back.

- Air radiation rates (microsieverts per hour)
- >19
 - 9.5–19
 - 1.0–9.5
 - 0.2–1.0
- Exclusion zones
- No public access
 - Restricted access
 - Free movement, no overnight stays

SLEEPING GIANT

In 2010, Japan was the world's third largest producer of nuclear power and had big plans to increase its output by 2030. The Fukushima disaster in 2011 has resulted in all of its reactors being turned off, and its expansion plans are on hold, although the current government may change that.

● Japan ○ World



48
434

Reactors operable February 2014

3
70

Reactors under construction February 2014

9
173

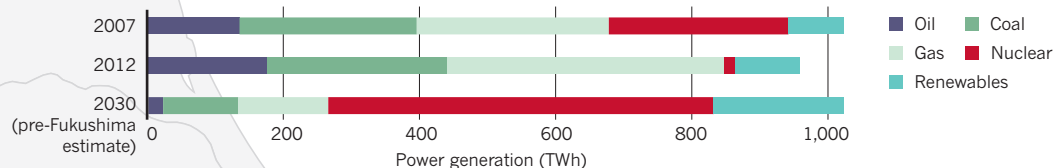
New-build reactors approved as of Feb 2014

3
310

New-build reactors proposed as of Feb 2014

IN THE MIX

Before the Fukushima disaster, Japan had planned to produce about 50% of its electricity from nuclear power by 2030.





Crystalline silica dust released during construction work can cause serious lung damage.

ENVIRONMENTAL HEALTH

Dust regulations trigger backlash

US agency's reassessment of silica exposure rules provokes conflict-of-interest row with senators.

BY DANIEL CRESSEY

Senate accusations of prejudice have forced a US government agency to defend its actions over a proposed tightening of regulations concerning industrial workers' exposure to deadly silica dust.

The row blew up late last year when the Occupational Safety and Health Administration (OSHA) began a public consultation on setting new limits for working with the dust, which is a major hazard for construction workers, causing serious lung disease. The agency ruffled feathers in the Senate when it asked that those submitting evidence should declare their funding sources.

Last November, a group of 16 senators wrote an open letter to OSHA criticizing the move for its implication that the agency might prejudice submissions. The consultation period closed on 11 February, and OSHA is now vigorously defending its request.

"What I'm doing here is essentially saying the information that we will base our standard on has to be of the highest integrity, and we have to do it in a transparent manner, and conflict-of-interest disclosure is an important

component of both of those," David Michaels, the head of OSHA, told *Nature*. "It would be surprising right now if a scientific journal didn't ask for that information."

Produced by tasks such as grinding concrete and sandblasting, used in the construction and other industries, crystalline silica dust can cause silicosis — an incurable disease involving inflammation of the lungs — and lung cancer. The dust is thought to kill or disable thousands of people in the United States every year, but guidelines on working with it have not been updated for more than 40 years.

"Our current standard is antiquated," says Michaels. "There are literally millions of workers in the United States who are exposed to dangerous levels of silica."

The present rules generally advise limiting exposure to roughly 100 micrograms of crystalline silica per cubic metre of air, averaged over 8 hours. OSHA has proposed halving this limit. Workers would also have to be better protected, for example by dust being 'wetted down' and with the use of extraction fans. OSHA estimates that the new regulations will cost about US\$640 million a year, with employers picking up most of the tab, but the agency

believes that the rules will save up to 700 lives a year. US standards are also influential in other countries, some note, potentially saving many more workers' lives.

The proposals were published in the *Federal Register* last September, at the start of the consultation period. In a first for OSHA, those wishing to submit scientific evidence as part of their comments were requested — although not required — to provide information on the funding sources of the research, as well as any funding received by the commenters that could potentially be considered a conflict of interest.

The Associated General Contractors of America, an industry group based in Arlington, Virginia, called the proposals "significantly flawed" and "rife with errors and inaccurate data". And shortly after they were published, the group of senators, led by Lamar Alexander (Republican, Tennessee), a senior member of the Senate Committee on Health, Education, Labor, and Pensions, wrote to OSHA saying that they were "very concerned about OSHA's attempt to have commenters disclose their financial backers". They added that the request "raises questions" about whether OSHA would prejudice submissions on the basis of who was sending them.

"The chilling effect the financial disclosure could have seems counter to the idea of robust inclusion of a diverse set of ideas and views to inform the rule-making," Liz Wolgemuth, a spokeswoman for Alexander, told *Nature*.

But pharmacologist Lisa Bero of the University of California, San Francisco, says that her own research on similar rule-making processes for tobacco control found that scientists opposing rules were often funded by industry groups. She supports the new disclosure request. "The regulatory agencies have to be in a position to critically appraise the studies that come to them," she says.

There is also support for the new silica standard. Tee Guidotti, a physician in Washington DC and a member of the American Thoracic Society's Environmental Health Policy Committee, says that the scientific case for the proposed limit is "close to being bullet-proof". He adds that, if it is successful, it could provide a template for how OSHA deals with similar hazards, such as dust and radon.

But Susan Dudley, director of George Washington University's Regulatory Studies Center in Washington DC, which conducts independent research on regulatory proposals, argues that there has already been a drop in exposure to silica dust and its health effects in recent years. She supports a lower exposure limit, but believes evidence is weaker for some of the specific requirements proposed to reach it, such as dust wetting.

The viewpoints contained in the 1,600 or so comments received through the consultation will be discussed in public hearings starting on 18 March. It will probably be several years before a final rule is enacted. ■ **SEE EDITORIAL P8**

QAPHOTOS.COM/LAWY

DIAGNOSTICS

Prenatal-screening companies expand scope of DNA tests

But the increasingly accurate analyses carry the ethical dilemma of uncertain outcomes.

BY ERIKA CHECK HAYDEN

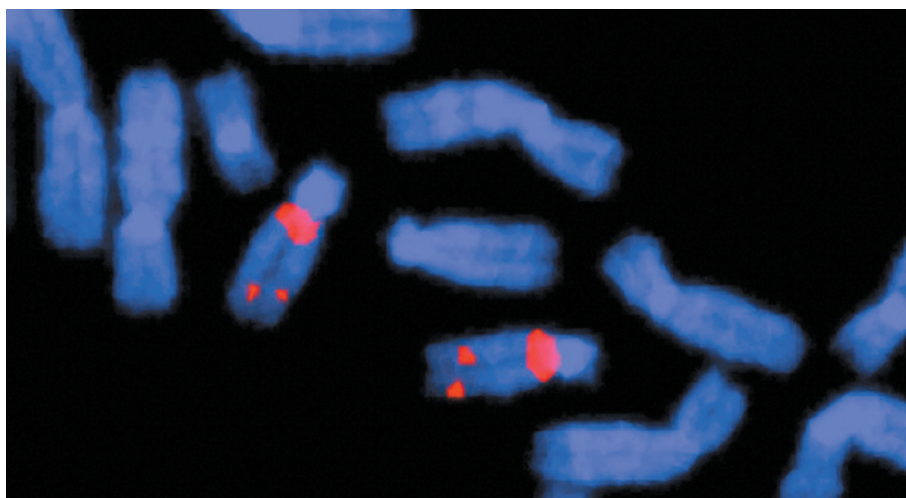
When the small piece of chromosome 22 associated with DiGeorge syndrome is missing, this can cause developmental abnormalities such as speech and language delays, cleft palate, a faulty immune system and heart defects. Or an apparently healthy baby can be born.

With non-invasive genetic screening on the rise, companies are racing to add chromosomal abnormalities such as that associated with DiGeorge syndrome to the list of disorders for which they can test. But because the effects of these genetic anomalies are sometimes uncertain, the test results may prompt difficult decisions for families who receive such diagnoses.

The tests are a welcome development, says Philippa Brice, communications director of the non-profit PHG Foundation in Cambridge, UK, which promotes the responsible use of biomedical science. But she adds, “there is a risk that women may receive early in pregnancy a result they do not fully understand”.

The issue is becoming more common as genetic-screening companies demonstrate the accuracy of the tests — and press their case that all pregnant women should be screened. On 26 February, scientists at Illumina, a genome-sequencing company headquartered in San Diego, California, announced the results of a study comparing its genetic test, called *verifi*, with the standard biochemical screening routinely offered to pregnant women (D. W. Bianchi *et al.* *N. Engl. J. Med.* <http://doi.org/rqn>; 2014). The genetic test checks directly for abnormalities in fetal DNA that crosses the placenta and enters the mother’s bloodstream, whereas the biochemical screen estimates the risk of chromosomal disorders indirectly, by measuring proteins and hormones in the mother’s blood that are linked with higher rates of birth defects. If either screen suggests the presence of a defect, it is usually confirmed by an invasive test, such as an amniocentesis.

The study, which was funded by Illumina, screened more than 1,900 US women for chromosomal abnormalities, including Down’s syndrome, using both the biochemical and genetic methods. The *verifi* test had a false-positive rate of 0.3%, about one-tenth of that associated with biochemical screening. The finding is crucial because most physicians’ groups recommend that screening be used only in women at



DNA deletions in chromosome 22 (indicated by pink markers) are associated with DiGeorge syndrome.

high risk of carrying a fetus with chromosomal disorders — mainly women aged over 35 — to minimize the chance that a false positive would lead to an invasive test. The study examined women with an average age of just under 30.

With such a low false-positive rate, the genetic screening should be offered to women of all ages, says Anthony Gregg, a geneticist and obstetrician at the University of Florida in Gainesville who was the lead author of genetic-screening guidelines published last April by the American College of Medical Genetics and Genomics in Bethesda, Maryland.

Increased acceptance could expand the already growing market for genetic screen-

“The companies are all trying to one-up each other.”

ing. Sequenom, a sequencing company based in San Diego, said in January that it performed 148,500 of its MaterniT21 tests in 2013, up 140% from 2012. And in November, Ariosa of San Jose, California, said that it had performed its Harmony test on 150,000 women since launching in May 2012. Most tests are performed in the United States, although the companies are preparing to roll them out globally.

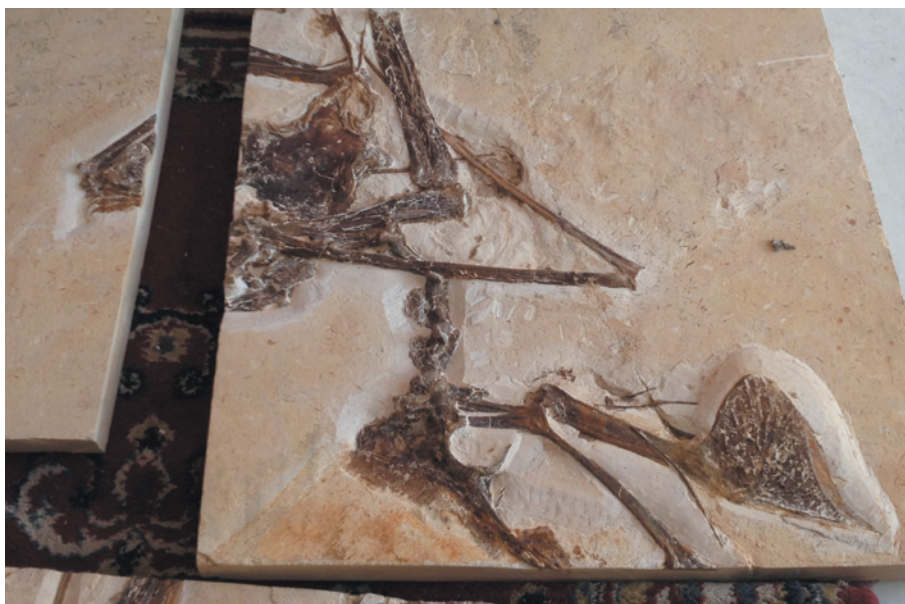
Companies are trying to stand out by expanding the number of conditions their tests check for. Microdeletions and microduplications — genetic defects that can be millions of base pairs long — are seen as the first step, although some companies have designs to sequence the entire fetal

genome. “The companies are all trying to one-up each other,” says Wendy Chung, a paediatrician and geneticist at Columbia University in New York City. “It’s not clear where the number will stop.”

Typically, genetic and biochemical screening tests have scanned only for missing or extra copies of whole chromosomes, which are known to result in clear outcomes. An extra copy of chromosome 21, for instance, causes Down’s syndrome. But last October, Sequenom began looking for microdeletions within chromosomes with its MaterniT21PLUS test. Natera, based in San Carlos, California, added tests for microdeletions to its Panorama test last month. And Illumina says that it will follow suit and expand the capabilities of its *verifi* test in the next few months. “Nobody in the twenty-first century should have a pregnancy without being screened for these microdeletions,” says Matthew Rabinowitz, Natera’s chief executive.

Routine checking for microdeletions could enable more families to prepare for children with special needs. But because such microdeletions are smaller than whole-chromosome abnormalities, they affect fewer genes, and the physical or mental effects are not always predictable. Chung, for instance, knows of many adults who found out that they had DiGeorge syndrome only after their children were diagnosed and they decided to get tested themselves. She says that as tests expand their scope and offer less-certain predictions, more families will face tough choices about their unborn children. ■

WESSEX REG. GENETICS CENTRE, WELLCOME IMAGES



Rare fossils, such as this pterosaur from Brazil's Araripe Basin, are regularly seized by authorities.

PALAEOLOGY

Brazil clamps down on illegal fossil trade

Debate rages as collectors accused of exporting specimens face up to 20 years in jail.

BY ELIZABETH GIBNEY

Thirteen people are scheduled to go on trial in Brazil for smuggling fossils out of the country, apparently to private collectors and to museums in Germany and the United Kingdom.

The arrests, which followed the seizure of around 1,000 fossils by French authorities in October, are part of a growing effort by Brazilian authorities to crack down on illegal fossil trading. Although the police have not released the names of the accused, authorities say that one, a German national, works for palaeontological museums in the United States, Europe and China. If convicted, some of the individuals face a maximum of 20 years in jail.

The case, which is still awaiting a trial date, has reopened the debate about the fossil trade, and whether science in developing countries benefits from protectionist policies.

In Brazil, it is illegal for anyone not affiliated with a national or state research institution to dig for fossils without authorization from the Brazilian National Department of Mineral Production (DNPM). The law dates back to 1942, but official permits are rarely given out. Brazil's constitution also says that

fossils found in the country are the property of the state, making their sale or export without permission illegal. "Despite this, people are taking fossils out of the country in tonnes per month," says Max Langer, president of the Brazilian Society of Palaeontology.

The fossil-rich Araripe Basin in northeastern Brazil has long been a source of black-market specimens. Langer says that he is not aware of the DNPM ever granting a licence to foreign collectors there, meaning that most Araripe fossils now abroad were obtained illegally.

Many Brazilian palaeontologists broadly support the law. They argue that keeping fossils in the country preserves cultural heritage and helps to develop Brazil's palaeontological expertise. Illegal collecting also damages specimens and prevents the acquisition of contextual field data, some say.

But the crackdown is worrying some palaeontologists outside Brazil, who say that the laws hamper science. "Scientists who just want to go about doing science are frustrated," says David Martill, a palaeontologist at the University of Portsmouth, UK.

The enforcement has been driven by Brazil's growing economy and strength in palaeontology, as well as an increasing public awareness

of fossil sales and loss of cultural heritage, says Alexander Kellner, a palaeontologist at the National Museum of Brazil, part of the Federal University of Rio de Janeiro. Authorities, for example, were quick to act last month when the online auction site eBay listed a Brazilian pterosaur for sale at US\$262,000. The auction, by French company Geofossiles, closed without sale, and Brazilian and international police are reportedly investigating the case. Brazil could ask France to repatriate the fossil if authorities can prove that it was found in the country.

Researchers say that the loss of fossils has seriously harmed Brazilian science. For instance, in findings presented at the 7th Brazilian Symposium of Vertebrate Palaeontology in 2010, Marcos Sales, a palaeontologist at the Federal University of Rio Grande do Sul in Porto Alegre, found that 14 of the 24 reference specimens, or holotypes, of pterosaur fossils collected in the Araripe Basin were housed by foreign institutions.

The issue is not unique to Brazil. Other countries that list fossils as national property and prohibit their export, such as China, Mongolia and Morocco, are increasingly enforcing their own laws. In 2012, for example, Mongolia successfully fought for the return of a 7.3-metre-long tarbosaur fossil, which had been illegally exported to the United States and sold there for more than \$1 million.

But critics such as Martill argue that criminalizing the fossil trade merely sends it underground. Crackdowns cut the supply line because informal diggers, such as quarry workers, fear arrest. Knowing "dodgy" people is the only way to get samples, says Martill, because the DNPM ignores requests to dig. He adds that palaeontology needs commercial trade because without it there is little excavation.

David Unwin, a palaeobiologist at the University of Leicester, UK, says that certain fossils are so abundant that there is no harm done if some are taken abroad. He also notes that the cost of preparing and displaying fossils is very high. "If a country is relatively poor and cannot afford those kinds of things, it is perhaps better that those fossils go somewhere else," he says. Langer calls this a "very first-world view".

Kellner admits that the Brazilian law is not perfect. He was a victim of it in 2012, when he was arrested while transporting fossils within Brazil, although the charges were dropped. He still supports the law, but says that the government's crackdown must be accompanied by a drive to give Brazilian palaeontologists the funding to collect fossils. Others say that the DNPM lacks the resources to enforce the law.

Langer feels that fossils must be kept in the country to help to improve Brazilian science. And he expects fellow researchers to hold Brazil's laws in higher regard than the private collectors who also fuel the trade. "Foreign researchers know it's illegal. They could work here in Brazil with Brazilian partners, but they choose not to," he says. ■

RESEARCH MANAGEMENT

Budget woes hinder clinical-trials network

Costly cancer trials a challenge for revamped US programme.

BY HEIDI LEDFORD

Following years of criticism, a nationwide US cancer clinical-trials programme came to an end on 1 March. But many fear that the programme replacing it — although designed to be nimbler and more focused — will be unable to conduct the types of trial needed in an era of expensive, personalized cancer treatments.

“We’ve got an exciting new structure and great science,” says Monica Bertagnolli, an oncologist at the Dana-Farber/Brigham and Women’s Cancer Center in Boston, Massachusetts, who serves on the new National Clinical Trials Network (NCTN). “But the severe budgetary constraints are forcing us to really limit what studies are being done.”

For more than half a century, the Clinical Trials Cooperative Group Program, run by the US National Cancer Institute (NCI) in Bethesda, Maryland, has been the main government-funded mechanism for carrying out trials that drug companies avoid because the results are unlikely to boost profits. The programme has had notable successes: it helped to boost the cure rate of childhood cancers from less than 10% to almost 80%; determined that the drug tamoxifen could halve the incidence of breast cancer in women at high risk of the disease; and found effective alternatives to radical mastectomies for early-stage breast cancers.

But the cooperative group programme has also been criticized for its inefficiency at a time when cancer research is advancing rapidly. Tumours have been divided into subtypes on the basis of their genetic mutations, and there is growing interest in testing tailor-made therapies. To achieve that, many patients must be screened to find those with the right cancer types.

In theory, the cooperative group programme, which included more than 3,000 institutions and some 14,000 investigators, could have provided the power and reach needed to find those patients. But a 2010 report by the US Institute of Medicine pronounced it unfit for the challenge. The institute found that many trials took so long to launch that, by the time they did, they no longer reflected cutting-edge science. Only about 60% of the programme’s trials were ever completed.

In response, over the past four years, the NCI has been readying a replacement: the

NCTN. Ten cooperative groups have been consolidated to five, and there is now a unified system for data management and ethical review. The NCTN also has strict deadlines for trial initiation. “If a study doesn’t meet those timelines, it’s dead,” says Bertagnolli. The time taken to start a large clinical trial has already dropped from two years to one — in line with the time it takes pharmaceutical companies to launch trials, says James Doroshow, deputy NCI director for clinical and translational research.

But the network’s budget, just like that of the old programme, falls far short of those of its industry counterparts. The NCI reimburses member institutions by about US\$2,000 per enrolled patient — much less than the

“What if a critical scientific opportunity arises that cannot be addressed?”

\$20,000 per patient that pharmaceutical companies provide, says Philip DiSaia, an oncologist at the University of California, Irvine, who was chair of a

gynaecological oncology cooperative group. As a result, some institutions have threatened to pull out of the network, says DiSaia, and others have capped the number of patients they will enrol in an NCTN trial.

At some clinical-trial sites, the NCI plans to boost reimbursement rates to \$4,000 per patient. But that, combined with a flat budget, means that the NCTN will admit only 17,000 patients in 2014, down from 21,000 admitted last year under the old system. And that could limit the kinds of trials the network can take on, says Robert Comis, co-chair of an NCTN group and an oncologist at Drexel University in Philadelphia, Pennsylvania. “What if a critical scientific opportunity arises that cannot be addressed because of the quotas?” ■

CORRECTION

In the News Feature ‘Smart enough to die?’ (*Nature* **506**, 284–286; 2014), it was the Virginia court system, not the US Supreme Court, that changed Daryl Atkins’s sentence to life imprisonment. The Supreme Court also did not mandate that rulings on intellectual disability in death-penalty cases should follow the standards of the American Association on Mental Retardation.



THE *SINS* OF THE FATHER

The roots of inheritance may extend beyond the genome, but the mechanisms remain a puzzle.

W

hen Brian Dias became a father last October, he was, like any new parent, mindful of the enormous responsibility that lay before him. From that moment on, every choice he made could affect his newborn son's physical and psychological development. But, unlike most new parents, Dias was also aware of the influence of his past experiences — not to mention those of his parents, his grandparents and beyond.

Where one's ancestors lived, or how much they valued education, can clearly have effects that pass down through the generations. But what about the legacy of their health: whether they smoked, endured famine or fought in a war?

As a postdoc in Kerry Ressler's laboratory

BY VIRGINIA HUGHES

at Emory University in Atlanta, Georgia, Dias had spent much of the two years before his son's birth studying these kinds of questions in mice. Specifically, he looked at how fear associated with a particular smell affects the animals and leaves an imprint on the brains of their descendants.

Dias had been exposing male mice to acetophenone — a chemical with a sweet, almond-like smell — and then giving them a mild foot shock. After being exposed to this treatment five times a day for three days, the mice became reliably fearful, freezing in the presence of acetophenone even when they received no shock.

Ten days later, Dias allowed the mice to mate with unexposed females. When their young grew up, many of the animals were more

sensitive to acetophenone than to other odours, and more likely to be startled by an unexpected noise during exposure to the smell. Their offspring — the 'grandchildren' of the mice trained to fear the smell — were also jumpier in the presence of acetophenone. What's more, all three generations had larger-than-normal 'M71 glomeruli', structures where acetophenone-sensitive neurons in the nose connect with neurons in the olfactory bulb. In the January issue of *Nature Neuroscience*¹, Dias and Ressler suggested that this hereditary transmission of environmental information was the result of epigenetics — chemical changes to the genome that affect how DNA is packaged and expressed without altering its sequence.

Biologists first observed this 'transgenerational epigenetic inheritance' in plants. Tomatoes, for example, pass along chemical markings that control an important ripening

gene². But, over the past few years, evidence has been accumulating that the phenomenon occurs in rodents and humans as well. The subject remains controversial, in part because it harks back to the discredited theories of Jean-Baptiste Lamarck, a nineteenth-century French biologist who proposed that organisms pass down acquired traits to future generations. To many modern biologists, that's "scary-sounding," says Oliver Rando, a molecular biologist at the University of Massachusetts Medical School in Worcester, whose work suggests that such inheritance does indeed happen in animals³. If it is true, he says, "Why hasn't this been obvious to all the brilliant researchers in the past hundred years of genetics?"

One reason why many remain sceptical is that the mechanism by which such inheritance might work is mysterious. Explaining it will require a deep dive into reproductive biology to demonstrate how the relevant signals might be formed in the germ line, the cells that develop into sperm and eggs and carry on, at a minimum, a person's genetic legacy.

A mother might pass on effects of environmental exposures to a fetus during pregnancy. So, to study the phenomenon of transgenerational epigenetics cleanly, biologists are focusing on fathers, and have been looking at how sperm might gain and lose epigenetic marks. "In the past two to three years there's been a lot of new information," says Michelle Lane, a reproductive biologist at the University of Adelaide in Australia. But proposals for how it all works are themselves embryonic. "It's a huge black box," Lane says.

MONSTER PLANTS AND OBESE CHILDREN

The epigenetics revolution hit in the early 2000s, when scientists began reporting that environmental factors — everything from neglectful mothering and child abuse to a high-fat diet and air pollution — can influence the addition or removal of chemical tags on DNA that turn genes on and off. This idea of an environmentally responsive genome still stirs debate (see *Nature* 467, 146–148; 2010). But the notion that epigenetic marks are transmitted across generations is even more provocative.

Swedish botanist Carl Linnaeus was among the first to spot changes resulting from this phenomenon. In the 1740s, he received a plant specimen that looked very similar to common toadflax (*Linaria vulgaris*), but with very different flowers. Linnaeus was shocked because this challenged his theory that plant species could be categorized by the structure of their flowers. "This is certainly no less remarkable," he wrote, "than if a cow were to give birth to a calf with a wolf's head." He named the plant *Peloria*, after the Greek word for 'monster'.

In the 1990s, plant biologist Enrico Coen at the John Innes Centre in Norwich, UK, found

that in the monster plants, methyl groups litter a gene involved in flower structure called *Lcyc*, completely shutting it down. (DNA methylation usually turns genes off.) Coen's team also showed that these methyl marks pass through seeds to later generations⁴.

The public first started to take notice in the

"WHY HASN'T THIS BEEN OBVIOUS TO ALL THE BRILLIANT RESEARCHERS IN THE PAST HUNDRED YEARS OF GENETICS?"

mid-2000s, after large epidemiological investigations in Europe began to show transgenerational effects in humans. One study of Swedish historical records showed that men who had experienced famine before puberty were less likely to have grandsons with heart disease or diabetes than men who had plenty to eat⁵. Similar work with children in Britain reported in 2005 that fathers who had started smoking before the age of 11 had an increased risk of having boys of above average weight⁶.

But many scientists remained sceptical. Epidemiological studies are often messy, and it is impossible to rule out all confounding variables. In the past few years, however, several studies in rodents have supported these observations and begun to attribute the transmission of various traits to changes in sperm.

SPERM SIGNATURES

Male rats fed a high-fat diet, for example, beget daughters with abnormal DNA methylation in the pancreas⁷. Male mice fed a low-protein diet have offspring with altered liver expression of cholesterol genes³. And male mice with pre-diabetes have abnormal sperm methylation, and pass on an increased risk of diabetes to the next two generations⁸.

"We and many other people have now shown these paternal effects," says Rando, who led the low-protein study. "And we're all having a hell of a time figuring out how they work."

The animal studies have triggered some strong debate. The most controversial results have come out of Michael Skinner's lab at Washington State University in Pullman. Skinner's team exposed pregnant rats to large doses of pesticides and fungicides, which led to organ damage in their adult offspring. The sperm of male offspring showed changes in DNA methylation that persisted for at least four generations⁹.

But at least two groups failed to replicate the data, and in 2010, federal investigators found

that one of Skinner's postdocs had fabricated data for a related paper, which the authors had retracted in 2009. Skinner says that some teams have replicated his results, and that those who have not were using inappropriate protocols. Last year, his own team reported successfully reproducing the results of the retracted paper¹⁰.

METHYLATION MECHANISM

Explaining how transgenerational epigenetics works has been difficult in part because most studies track outcomes — such as changes in glucose, cholesterol and fertility — that can be affected by a range of factors, making it tricky to tease out cause and effect. By contrast, Dias and Ressler's work with acetophenone takes advantage of specific biology: the chemical binds to a particular receptor in the nose that is encoded by a single gene, dubbed *Olf151*. "This is the massive pro of their study," Rando says.

Dias and Ressler do not claim to understand exactly what is going on, but they do have a working hypothesis. Somehow, the information about the frightening smell gets into a mouse's testes and results in lower methylation of the *Olf151* gene in sperm DNA. The researchers even ran experiments using *in vitro* fertilization to make sure that the father was not in some way passing on a fear of acetophenone through interactions with the mother. The epigenetic tweak in the sperm is perpetuated in the offspring's DNA, leading to increased expression of the receptor in the animals' noses and, ultimately, enhanced sensitivity to the smell.

But the chain of causation is loose. "There are a lot of disconnects there," says William Kelly, a developmental geneticist at Emory. "It's not beyond the realm of possibility or plausibility. It's just right now we don't know enough about how information is transferred between generations."

The first question is how the effects of environmental exposure become embedded in an animal's germ cells — in this case, the mouse's sperm. Germ cells have been shown to express olfactory receptors¹¹. So it is possible that *Olf151* receptors in sperm respond to odorant molecules in the bloodstream and then change the methylation of the corresponding gene in sperm DNA.

Alternatively, after being exposed to the odour and the pain, a mouse might produce RNA molecules — perhaps in the brain — that make their way into the bloodstream and then selectively target the *Olf151* gene in sperm. Many studies in plants have hinted at this sort of systemic RNA shuttling. RNA molecules expressed in a plant's leaf, for example, can travel through its vascular system to many of its other tissues and affect gene expression¹².

But creating an epigenetic mark in the sperm is only the first step. To pass down through multiple generations, the signal needs to survive multiple rounds of rigorous

epigenetic reprogramming. In mammals, the first of these happens just hours after conception, when most methylation is stripped from sperm DNA in the single-celled embryo. Then, as the embryo develops and divides, and cells begin to differentiate into various tissue types, methylation is gradually re-established. But even if some signal from the father were to survive this process, the embryo's own primordial germ cells, those that eventually become its sperm or eggs, undergo a second round of epigenetic scrubbing (See 'Without a trace').

Some genes manage to escape these periods of major reprogramming. The best example is genes that are imprinted — whereby one copy from the mother or father is robustly methylated and effectively silenced. These silencing marks crop up in the egg or sperm and are retained in the embryo.

About 100 genes are known to be imprinted, but some non-imprinted genes may also escape the scrubbing through a similar mechanism. "There is a growing consensus that there are more regions than previously thought that escape reprogramming in sperm," says Sarah Kimmins, an epigeneticist at McGill University in Montreal, Canada. "Why this is, and how, is not yet known, although studying imprinted genes may reveal clues."

Then again, even if *Olfr151* does escape reprogramming, it is hard to explain how that could lead to a noticeable difference in the behaviour of fully formed offspring. Dias and Ressler reported that in sperm samples from mice trained to fear acetophenone, about 86 out of every 100 sperm show *Olfr151* methylation, whereas in mice trained to fear a different odour it is about 95 out of every 100. This difference is statistically significant, but fairly small. And yet the behavioural effects in the second generation were robust: about half of the acetophenone-trained animals' offspring showed increased sensitivity to the odour.

'SOMETHING GOOFBALL'

Although many are scratching their heads over the holes in the proposed mechanism, few are suggesting that the underlying phenomenon is a fairy tale. "Impossible things are happening every day," says Kelly, quoting a line from Rodgers and Hammerstein's *Cinderella*.

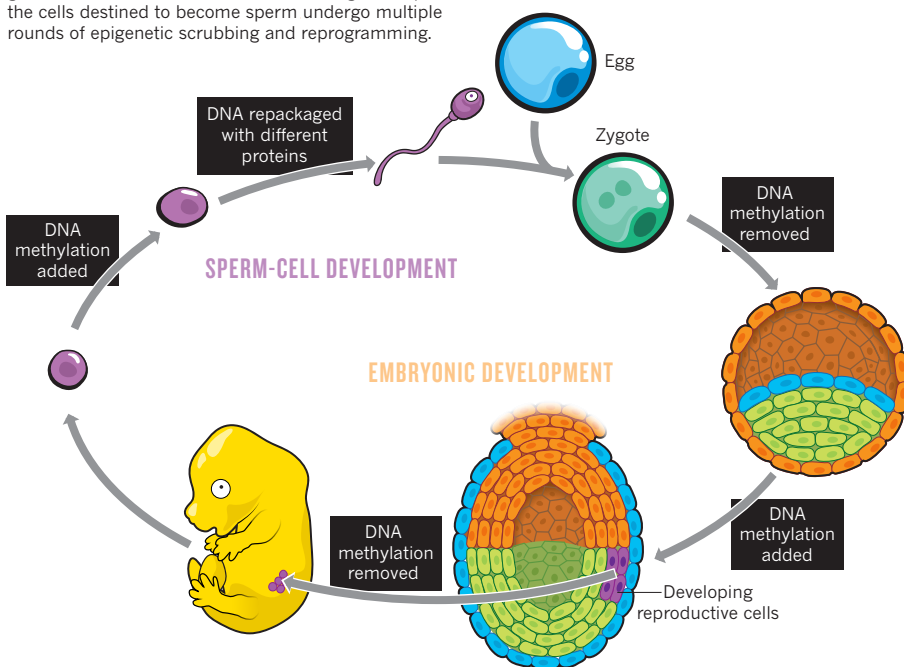
It is possible, for example, that the DNA-methylation tweaks reported in the odour study are simply a by-product of an altogether different mechanism.

One route might be chemical marks on histones, the proteins around which DNA wraps. Acetyl and methyl groups can attach to histones and affect the expression of nearby DNA. But during sperm-cell formation, DNA is stripped of most of its histones (and their attendant marks) and wraps instead around protamines, which pack it more tightly.

Nevertheless, about 10% of human histones — and about 1% of mouse ones — are retained. These sites might carry information from one

WITHOUT A TRACE

Researchers are struggling to understand how epigenetic marks, such as DNA methylation, could pass from one generation to the next in mammals. During development, the cells destined to become sperm undergo multiple rounds of epigenetic scrubbing and reprogramming.



generation to the next. In 2011, researchers reported that, in nematode worms, certain histone marks correlate with long life and can be passed down through several generations¹³. And last December, Kimmins and her colleagues showed that feeding male mice a diet low in folate — a nutrient that provides the raw materials for methylation — led to significantly reduced methylation of histone proteins in the animals' sperm and more birth defects in their offspring¹⁴.

Still other studies point to a mechanism involving short RNA molecules latching on to DNA and affecting gene expression. Twenty-eight microRNAs are expressed differently in the sperm of men who do and do not smoke, according to a study reported in 2012 (ref. 15). And these RNA patterns may persist through multiple generations. Last year, Lane's group found that obese male mice show abnormal expression of 11 microRNAs in their sperm — and that they pass on insulin resistance to the next two generations¹⁶.

Then there is the possibility that the mechanism is, as Rando puts it, "something goofball". That might be prions — misfolded proteins that act as infectious agents — which have been shown to transmit heritable traits in budding yeast (see *Nature* 482, 294–296; 2012). Or it could be something in semen besides sperm. Researchers reported in January¹⁷ that mice born of fathers lacking seminal vesicles are fatter and have more metabolic problems than controls, suggesting that molecules in seminal fluid influence gene expression in sperm and the female reproductive tract.

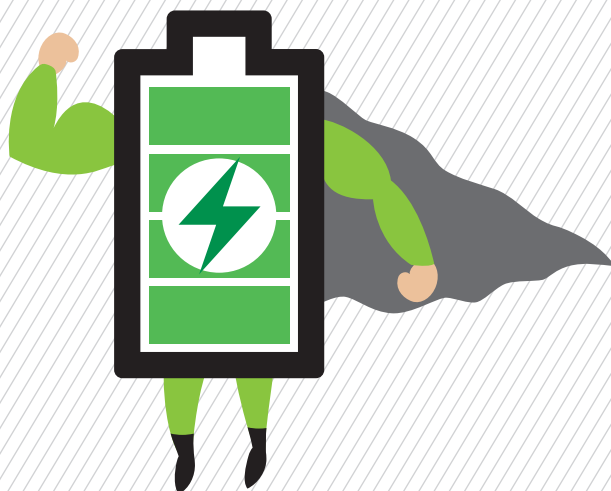
If the mechanism involves DNA

methylation, histones or RNA, the field is likely to make great progress in the next few years, Rando predicts. "But if it's something completely novel," he says, "Maybe it will take decades to figure out."

Dias has his fingers crossed for the former. He is going to Boston, Massachusetts, in April for a Keystone meeting on epigenetic inheritance, to get a sense of the most promising mechanistic avenues to follow. "If science has taught me anything," he says, "it is to not discount the myriad ways of becoming and being." ■

Virginia Hughes is a freelance journalist in New York City.

1. Dias, B. G. & Ressler, K. J. *Nature Neurosci.* **17**, 89–96 (2014).
2. Manning, K. et al. *Nature Genet.* **38**, 948–952 (2006).
3. Carone, B. R. et al. *Cell* **143**, 1084–1096 (2010).
4. Cubas, P., Vincent, C. & Coen, E. *Nature* **401**, 157–161 (1999).
5. Kaati, G., Bygren, L. O. & Edvinsson, S. *Eur. J. Hum. Genet.* **10**, 682–688 (2002).
6. Pembrey, M. E. et al. *Eur. J. Hum. Genet.* **14**, 159–166 (2006).
7. Ng, S.-F. et al. *Nature* **467**, 963–966 (2010).
8. Wei, Y. et al. *Proc. Natl Acad. Sci. USA* **111**, 1873–1878 (2014).
9. Anway, M. D., Cupp, A. S., Uzumcu, M. & Skinner, M. K. *Science* **308**, 1466–1469 (2005).
10. Skinner, M. K., Haque, C. G.-B., Nilsson, E., Bhandari, R. & McCarrey, J. R. *PLoS ONE* **8**, e66318 (2013).
11. Goto, T., Salpekar, A. & Monk, M. *Mol. Hum. Reprod.* **7**, 553–558 (2001).
12. Dunoyer, P. et al. *Science* **328**, 912–916 (2010).
13. Greer, E. L. et al. *Nature* **479**, 365–371 (2011).
14. Lambrot, R. et al. *Nature Commun.* **4**, 2889 (2013).
15. Marczylo, E. L., Amoako, A. A., Konje, J. C., Gant, T. W. & Marczylo, T. H. *Epigenetics* **7**, 432–439 (2012).
16. Fullston, T. et al. *FASEB J.* **27**, 4226–4243 (2013).
17. Bromfield, J. J. et al. *Proc. Natl Acad. Sci. USA* **111**, 2200–2205 (2014).



A BETTER BATTERY

Chemists are reinventing rechargeable cells to drive down costs and boost capacity.

BY RICHARD VAN NOORDEN

The mobile world depends on lithium-ion batteries — today's ultimate rechargeable energy store. Last year, consumers bought five billion Li-ion cells to supply power-hungry laptops, cameras, mobile phones and electric cars. "It is the best battery technology anyone has ever seen," says George Crabtree, director of the US Joint Center for Energy Storage Research (JCESR), which is based at the Argonne National Laboratory near Chicago, Illinois. But Crabtree wants to do much, much better.

Modern Li-ion batteries hold more than twice as much energy by weight as the first commercial versions sold by Sony in 1991 — and are ten times cheaper. But they are nearing their limit. Most researchers think that improvements to Li-ion cells can squeeze in at most 30% more energy by weight (see 'Powering up'). That means that Li-ion cells will never give electric cars the 800-kilometre range of a petrol tank, or supply power-hungry smartphones with many days of juice.

In 2012, the JCESR hub won US\$120 million from the US Department of Energy to take a leap beyond Li-ion technology. Its stated goal was to make cells that, when scaled up to the sort of commercial battery packs used in electric cars, would be five times more energy dense than the standard of the day, and five

times cheaper, in just five years. That means hitting a target of 400 watt-hours per kilogram (Wh kg^{-1}) by 2017.

Crabtree calls the goal "very aggressive"; veteran battery researcher Jeff Dahn at Dalhousie University in Halifax, Canada, calls it "impossible". The energy density of rechargeable batteries has risen only sixfold since the early lead-nickel rechargeables of the 1900s. But, says Dahn, the JCESR's target focuses attention on technologies that will be crucial in helping the world to switch to renewable energy sources — storing up solar energy for nighttime or a rainy day, for example. And the US hub is far from alone. Many research teams and companies in Asia, the Americas and Europe are looking beyond Li-ion, and are pursuing strategies that may topple it from its throne.

LOSE THE DEAD WEIGHT

Chemical engineer Elton Cairns suspected he had tamed a promising-but-wild battery chemistry early last year, when his coin-sized cells were still going strong even after a few months of continual draining and recharging. By July, his cells at the Lawrence Berkeley National Laboratory in Berkeley, California, had cycled 1,500 times and had lost only half of their capacity¹ — a performance roughly on a par with the best Li-ion batteries.

His batteries are based on lithium-sulphur (Li-S) technology, which uses extremely cheap materials and in theory can pack in five times more energy by weight than Li-ion (in practice, researchers suspect, it will probably be only twice as much). Li-S batteries were first posited 40 years ago, but researchers could not get them to survive past about 100 cycles. Now, many think that the devices are the technology closest to becoming a commercially viable successor to Li-ion.

One of Li-S's main advantages, says Cairns, is that it gets rid of the "dead weight" in a Li-ion battery. Inside a typical Li-ion cell, space is taken up by a layered graphite electrode that does little more than host lithium ions. These ions flow through a charge-carrying liquid electrolyte into a layered metal oxide electrode. As with all batteries, current is generated because electrons must flow around an outside circuit to balance the charges (see 'Radical redesigns'). To recharge the battery, a voltage is applied to reverse the electron flow, which also drives the lithium ions back.

In a Li-S battery, the graphite is replaced by a sliver of pure lithium metal that does double duty as both the electrode and the supplier of lithium ions: it shrinks as the battery runs, and reforms when the battery is recharged. And the metal oxide is replaced by cheaper,

lighter sulphur that can really pack the lithium in: each sulphur atom bonds to two lithium atoms, whereas it takes more than one metal atom to bond to just one lithium. All of that creates a distinct weight and cost advantage for Li–S technology.

But the reaction between lithium and sulphur causes a problem. As the battery is charged and discharged, soluble Li–S compounds can seep into the electrolyte, degrading the electrodes so that the battery loses charge and the cell gums up. To prevent this, Cairns uses tricks made possible by advances in nanotechnology and electrolyte chemistry — including adulterating his sulphur electrode with graphene oxide binders, and using specially designed electrolytes that do not dissolve lithium and sulphur so much. Cairns predicts that a commercial-sized cell could achieve an energy-density of around 500 Wh kg^{-1} . Other labs are reporting similar results, he says.

Some researchers doubt that the academic cheer will translate into commercial success. Laboratories often use low proportions of sulphur and lots of electrolyte, which is relatively easy to work with but does not create an energy-dense battery. Bumping up the sulphur and decreasing the electrolyte makes the cell more likely to gum up, says Steve Visco, who has spent more than 20 years working on Li–S at battery firm PolyPlus in Berkeley, just 5 kilometres west of Cairns' lab. Making a cheap commercial cell that works over a range of temperatures will also be hard, he says.

At least one company stands by Li–S's prospects: Oxis Energy in Abingdon, UK. It says it has run large cells for an impressive 900 cycles, at energy densities that match current Li-ion cells. Oxis is working with Lotus Engineering, headquartered in Ann Arbor, Michigan, on a project to reach 400 Wh kg^{-1} by 2016 for an electric vehicle.

PACK MORE PUNCH PER ION

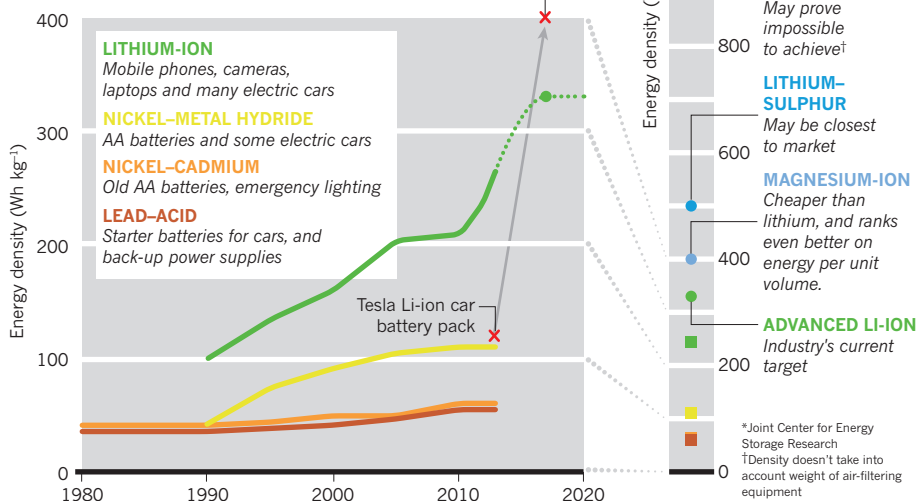
As the world's lightest metal, lithium provides a huge weight advantage. But some researchers argue that the next generation of cells should switch to heavier elements such as magnesium. Unlike lithium ions, which can carry only one electrical charge each, doubly charged magnesium ions shuttle two at a time — instantly multiplying the electrical energy that can be released for the same volume.

Magnesium comes with its own challenge, however: whereas lithium zips through electrolytes and electrodes, magnesium with its two charges moves as if through treacle.

Peter Chupas, a battery researcher at Argonne National Laboratory who is working with the JCESR, is shooting high-energy X-rays at magnesium in various electrolytes to investigate why it experiences so much drag. So far, he and his colleagues have found that magnesium exerts a strong pull on oxygen atoms in any surrounding solvent, attracting clusters of solvent molecules that make it bulkier. That kind of

POWERING UP

Portable rechargeable batteries tend to hit an energy-storage-per-weight limit. Lithium-ion technology has gone through several phases and types, but is also expected to reach a ceiling soon.



basic research is key to creating a better battery, but it is not usually done by industry, says Crabtree. “The typical R&D operation operates on trial and error, not fundamental research,” he says. This, he says, is where JCESR is bringing an advantage to the field.

Materials scientist Kristin Persson at Lawrence Berkeley is using a supercomputer to simulate the innards of possible new batteries, trying to find a combination of electrodes and electrolytes that will allow magnesium to pass through more easily. “Right now, we are crunching through around 2,000 different electrolytes,” she says.

Persson and Gerbrand Ceder, a materials

As companies jostle in secret, Persson continues to run through what she calls the “electrolyte genome”. The sifting-by-supercomputer approach could also help the search for batteries made with other multiple-charge-carrying (or ‘multivalent’) metals, such as aluminium and calcium. Ceder urges patience, pointing out that research into Li-ion battery chemistry has enjoyed a 40-year head start. “We have so little information about multivalent ions,” he says.

MAKE BATTERIES THAT BREATHE

Winfried Wilcke, who describes himself as “an extremely happy owner of a Tesla S” electric car, credits the vehicle with changing his mind about battery-research priorities.

Five years ago, Wilcke, who heads IBM's nanoscience and technology division in San Jose, California, launched a project to develop a car battery with an 800-kilometre range. At the start, he focused on the theoretical ultimate in energy-dense electrochemical storage: the oxidation of lithium with oxygen drawn from the air. Such ‘breathing’ batteries have a huge weight advantage over other types, because they do not have to carry around one of their main ingredients. A lithium–oxygen (Li–O) battery can, in theory, store energy as densely as a petrol engine — more than ten times better than today's car battery packs.

But after driving more than 22,000 kilometres in his electric roadster, Wilcke is happy with the 400-kilometre range that its battery already provides. The real problem, he says, is money: battery packs for electric cars cost more than $\$500 \text{ kWh}^{-1}$. “What's holding back the mass acceptance of electric cars is really the price rather than the energy density,” he says. So Wilcke now favours a cheaper breathing battery based on sodium. Theory predicts that sodium–oxygen (Na–O) batteries could provide only half

Five times more energy dense, and five times cheaper, in just five years: an “impossible” goal?

scientist at the Massachusetts Institute of Technology in Cambridge, founded a company to develop these higher-charge-carrying batteries. Pellion Technologies, based in Cambridge, is tight-lipped about its results; it has published only one paper about electrolytes². A spate of patents published in late 2013 hint that the company is developing more-open electrode structures to help the magnesium ions to flow. Major electronics firms such as Toyota, LG, Samsung and Hitachi are also working on such cells, releasing little information beyond occasional teasers.

the energy density of Li–O, but that is still five times better than Li-ion batteries. And sodium is cheaper than lithium, so Na–O might, Wilcke hopes, get closer to the \$100-kWh⁻¹ goal that the JCESR and others have set for affordability.

Wilcke's change of heart was undoubtedly influenced by the fact that many have given up hope on Li–O. Researchers who have tried to make it work over the past 20 years have wrestled with unwanted side reactions: carbon in the electrolyte and electrode material react with the lithium and oxygen to form lithium carbonate, so that in every cycle, some 5–10% of the battery capacity is lost. After 50 cycles or so, the battery suffocates. "The bottom line is that Li–O has zero chance for vehicles," says Stanley Whittingham at Binghamton University in New York, who invented the concept of Li-ion batteries in the 1970s and still focuses on squeezing the best performance out of them. Researchers hoping to resuscitate Li–O include Peter Bruce, a chemist at the University of St Andrews, UK. "We are closer to what's needed than we were a few years ago," he argues. But many consider it a lost cause.

Wilcke took an interest in the sodium breathing battery last year, following a surprising discovery by a team including Jürgen Janek and Philipp Adelhelm at the Justus-Liebig University of Giessen in Germany. They found that a Na–O battery recharges more efficiently than Li–O, without complicating side-reactions³. "We tried it and were pretty stunned," says Wilcke. Plus, he says, it works with cheap electrodes and electrolytes. Janek says that his team has now shown that its battery can work reversibly for at least 100 cycles — not bad for the early days of the technology. Chemicals giant BASF is now working with them.

Dahn, for one, is not convinced. Debate rages about whether breathing batteries will require heavy filtering equipment to extract oxygen from the air, which would cut down or even eliminate their energy-per-weight advantage. "Na–O is just the latest craze," says Dahn. But Wilcke is willing to bet otherwise.

GO BIG FOR THE GRID

Donald Sadoway's vision of the future battery looks like a smelting plant: he envisions crates the size of shipping containers, each holding 20 refrigerator-sized steel blocks containing litres of molten metals and salts heated to 500 °C.

Such batteries could never fit in a car, and

cannot beat Li-ion on measures such as energy stored per unit weight. But when it comes to storing energy for the electricity grid — or other non-portable applications — size does not matter. Instead of a small, light battery that packs a powerful punch, what people need is a battery that cheaply bottles and releases

test batteries in Hawaii and at a military base on Cape Cod, Massachusetts, this year, each supplying tens of kilowatt-hours.

Other research groups are pursuing less-radical flow batteries, in which the fuel consists of two liquids that pass ions to each other through a membrane. The liquids can be kept in tanks outside the battery and pumped in to flow past each other when needed, so it is possible to store larger amounts of energy indefinitely simply by using bigger tanks. But they do need pumps and valves, which Sadoway says will require maintenance.

Commercial flow batteries use vanadium ions in the liquid on both sides of the barrier. But vanadium and the membranes are expensive: the world's largest flow battery, installed at a wind farm in China, probably costs \$1,000 kWh⁻¹, estimates Huamin Zhang at the Chinese Academy of Sciences' Dalian Institute of Chemical Physics. "The cost of vanadium just kills you," says Michael Aziz, a materials scientist at Harvard University in Cambridge, Massachusetts.

In January this year, a team including Aziz announced⁴ that cheap organic chemicals called quinones could be used in a flow battery, partnered to a standard liquid electrode such as bromine. Aziz has cycled his system more than 100 times and it is still running strong. He hopes that he can get such batteries below the magic \$100 kWh⁻¹, but "this is a toy in a fume hood in a laboratory right now," he says. "There is no way to know the true cost until you are mass-producing it."

Crabtree calls the work "promising" and says that the JCESR is also looking at organic chemicals for flow batteries. Another option it is pursuing is to use liquid Li–S and solid lithium in a sort of half-flow battery.

"It's early days: people are looking at really oddball systems, and everyone's trying to figure out how to get the lifetime up and the costs down," says Dahn. The JCESR, for one, is hoping that basic research can fill in the gaps and make these technologies work. "The beyond-lithium-ion space is rich with opportunity," says Crabtree, "and mostly unexplored." ■

Richard Van Noorden is a senior reporter for *Nature* in London.

RADICAL REDESIGNS

Lithium-ion batteries are today's best choice for portable, rechargeable applications. Better batteries could be made by changing the electrodes, the electrolyte or the charge-carrying ions. Researchers are also pursuing other designs.

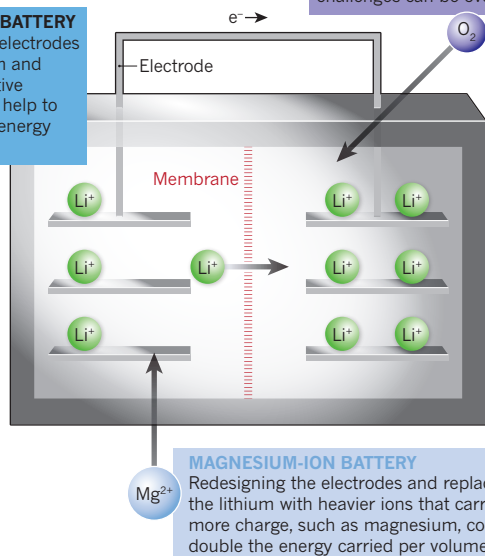
LITHIUM-ION BATTERY

A chemical energy gradient drives lithium ions through a membrane from a graphite electrode to a metal oxide one, causing electrons to flow around a closed circuit.

LITHIUM-OXYGEN

Batteries that pull in oxygen from the air could pack a serious punch, if major technical challenges can be overcome.

LI-SULPHUR BATTERY
Changing the electrodes to solid lithium and chemically active sulphur could help to pack in more energy per kilogram.



MAGNESIUM-ION BATTERY
Redesigning the electrodes and replacing the lithium with heavier ions that carry more charge, such as magnesium, could double the energy carried per volume.

small-to-large amounts of electricity without much maintenance. The JCESR wants such batteries to last for 7,000 cycles, or about 20 years.

"The field is wide open," says Ceder. Grid suppliers have used banks of cheap, old-fashioned lead–acid batteries, for example, or stacks of Li-ion. A dizzying array of other chemistries are in development, including zinc–air and sodium-ion. Most technologies are doing well to cost five times as much as the JCESR's \$100-kWh⁻¹ target.

Sadoway, a materials chemist at the Massachusetts Institute of Technology, is developing an alternative with two layers of molten metal as electrodes, separated by their different densities and by a layer of molten-salt electrolyte. The metal layers swell or shrink as ions pass between them, storing or releasing energy. Because everything is liquid, there is nothing that could crack after thousands of cycles, as solid electrodes might.

Crabtree, Dahn and other researchers worry about the energy needed to keep the components molten. But Sadoway says that the charging and discharging processes produce enough heat on their own. His company — Ambri in Marlborough, Massachusetts — plans to install

1. Song, M.-K., Zhang, Y. & Cairns, E. J. *Nano Lett.* **13**, 5891–5899 (2013).
2. Doe, R. E. *et al. Chem. Commun.* **50**, 243–245 (2014).
3. Hartmann, P. *et al. Nature Mater.* **12**, 228–232 (2013).
4. Huskinson, B. *et al. Nature* **505**, 195–198 (2014).

COMMENT

FOOD Eight ways to make ruminant farming more sustainable **p.32**

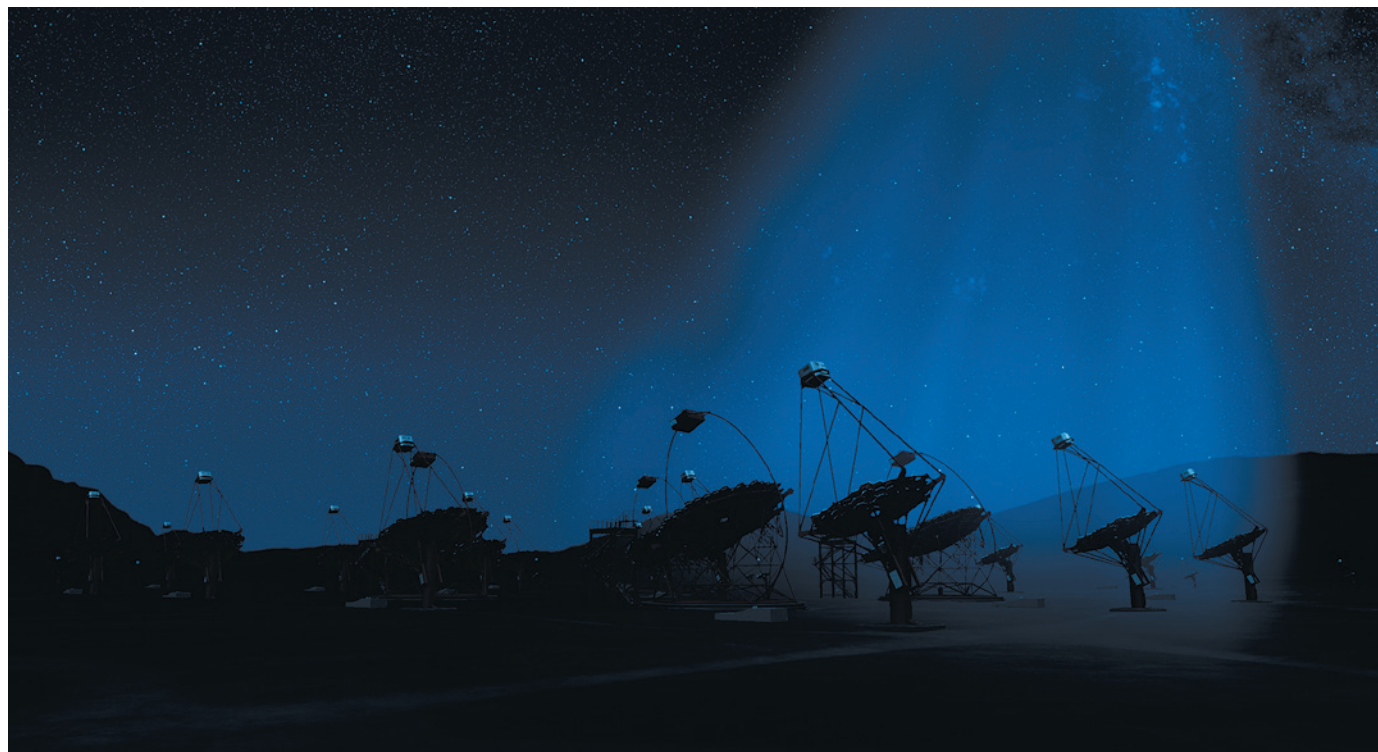
CRYPTOGRAPHY Chinese blockbuster novel on code-breaking reviewed **p.36**

HOMININS Human-origins exhibition probes history of immigration **p.38**



CONSERVATION Preserving concrete, from dams to Fallingwater **p.39**

DESIGN/MILDE SCIENCE COMM/EXOZET



The proposed Cherenkov Telescope Array (artist's impression) might detect light flashes from γ -rays produced when dark-matter particles interact.

Broaden the search for dark matter

Bold strategies are needed to identify the elusive particles that should make up most of the Universe's mass, say **Mario Livio** and **Joe Silk**.

Dark matter is living up to its name. In spite of decades of compelling evidence from astronomical observations showing the existence of matter that neither emits nor absorbs electromagnetic radiation, all attempts to detect dark matter's constituents have failed.

The presence of dark matter is inferred from its gravitational effects. Stars and gas clouds in galaxies and galaxies in clusters move faster than can be explained by the pull of visible matter alone. Light from distant objects may be distorted by the gravity

of intervening dark material. The pattern of large-scale structures across the Universe is largely dictated by dark matter. In fact, about 85% of the Universe's mass is dark, accounting for about one-quarter of the total cosmic energy budget.

Despite its ubiquity, the nature of dark matter eludes us. Negative results have flowed from searches for candidate particles to explain it. In 2013, the Large Underground Xenon (LUX) experiment — the most sensitive detector of its kind — in the Homestake Mine in Lead, South Dakota, reported no

signs of dark matter in its first three months of operation¹. The Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, has found no evidence for the existence of what some think are the most likely culprits: supersymmetric particles, theoretically predicted partners to the known elementary particles.

Is there light at the end of this dark tunnel? Possibly — but only if searches become bolder and broader. More varied particle types should be sought. Definitive tests need to be devised to rule out some classes of ►

► dark matter and some theories. If dark matter remains undiscovered in the next decade, then physicists will have to seriously reconsider alternative theories of gravity.

EXOTIC PARTICLES

We know a little about dark matter². Because it does not absorb light or interact with electromagnetic waves, the majority of it cannot be made of baryons — particles of ordinary matter, including protons and neutrons, which are composed of three quarks. And dark matter must lie beyond the standard model of particle physics to avoid upsetting Big Bang nucleosynthesis, the theory of which successfully predicts the observed abundances of light elements such as deuterium, helium and lithium arising from interactions in the early Universe.

The main constituents are expected to be weakly interacting massive particles (WIMPs)³. These particles have masses a few tens to thousands of times that of the proton. WIMPs interact among themselves and with ordinary matter gravitationally and through the weak force, but not electromagnetically or through the strong nuclear force. To explain the way in which galaxies form and cluster, dark-matter particles should be relatively slow moving, or 'cold'. If they were faster, and could move easily beyond the dimensions of a protogalaxy, many structures visible today would have been washed out.

What kind of particles might WIMPs be? Many physicists guess the most likely candidates are the lightest of the supersymmetric (SUSY) particles. Theories of the early Universe contain mathematical symmetries that allow for each known particle, such as the electron, photon or quark, to have a (yet undiscovered) massive partner. Just after the Big Bang, almost all of the heavy SUSY particles decayed or annihilated one another — but the lightest SUSY particles, unable to decay any further, were stable and survived.

The number of SUSY particles remaining depends on their masses and interaction strengths, which can be predicted from theory. The properties must be just right for the relic particle density to match that needed to explain the observed effects of dark matter. Light SUSY particles fit that bill, motivating a plethora of experiments to try to detect them. Yet no WIMPs have been seen.

Finding new particles is challenging. It took four decades to spot the Higgs boson. But through current and planned experiments we should be able to rule out several candidates for dark matter in the next decade. The goal is to detect the particles that constitute the massive halo of dark matter that surrounds the Milky Way as they pass through our detectors at rates of a few per second per square metre⁴. Because they interact so weakly, it takes a huge effort to capture them. If we are lucky, a 300-day run of LUX, scheduled to begin later

this year, could detect WIMPs. Or it may take bigger detectors — and more than one method (see 'What's the matter?').

Some estimate that detectors with 100 times more mass than LUX will be needed to see WIMPs at the rates expected. LUX ZEPLIN, a planned upgrade that could begin operation around 2019, would use 7 tonnes of liquid xenon compared with LUX's current 370 kilograms. We may yet require 100-tonne detectors, but that is the practical limit. Further sensitivity is thwarted by an irreducible neutrino background, mainly from supernovae, the Sun and cosmic rays hitting Earth's atmosphere.

Dark-matter particles might be created in colliders. The LHC is expected to resume

"Some theorists have even started to wonder whether dark matter exists."

operation in 2015 at an energy of 14 teraelectronvolts (TeV), twice the energy that led to the discovery of the Higgs boson. SUSY particles and signatures of departure from the standard model might be glimpsed. But the lack of any SUSY signals so far suggests that we may need much higher energies to see them. A 100-TeV collider, which many particle physicists support as the next step after the LHC, is slated to begin construction around 2020 and would be exciting for dark-matter searches⁵.

DIFFICULT TO DETECT

Other dark-matter experiments have made intriguing detections, the interpretations of which are still debated. The DAMA/LIBRA experiment, and its predecessor, at Gran Sasso National Laboratory in Italy, have been looking for changes in the flux of dark-matter particles hitting Earth for 14 years. As Earth orbits the Sun and as both travel through the Milky Way, the velocities of the two bodies combine. For half the year, their

velocities are in the same direction; for the other half, they are in opposition. This produces an annual modulation in the rate at which the dark-matter particles, travelling in random orbits, fall on the detector.

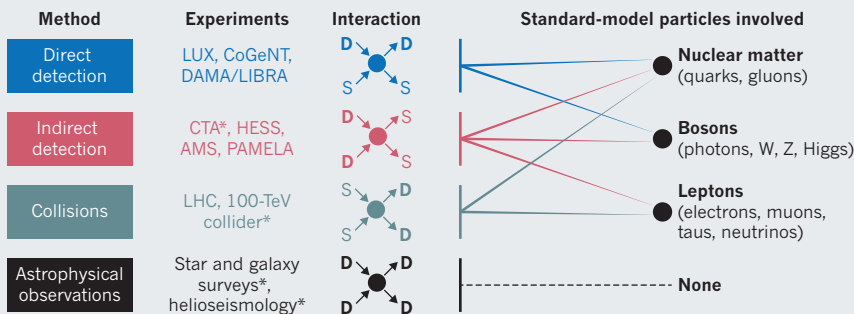
In 2013, the DAMA/LIBRA team reported⁶ such a variation with an accumulated high statistical significance (more than nine sigma). Similar cycles, consistent in phase but with an amplitude larger than expected, were seen (with low statistical significance) independently earlier this year in just over three years of data from the CoGeNT (Coherent Germanium Neutrino Technology) Dark Matter Experiment⁷ in Soudan, Minnesota. Yet most physicists query whether the DAMA/LIBRA results really are because of WIMPs and not some other annual phenomenon, such as neutrons leached from the rocks surrounding the underground experiment in response to seasonal temperature variations.

Indirect attempts to detect dark matter⁸ have been equally inconclusive. The Alpha Magnetic Spectrometer (AMS-02) on the International Space Station reported⁹ last year an excess of positrons in the cosmic-ray spectrum up to 350 gigaelectronvolts (GeV), consistent with being produced by dark-matter particles colliding and annihilating. These results strengthened similar reports from the Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics (PAMELA) satellite. But the positrons could have other origins, such as winds from pulsars (rapidly rotating neutron stars). Observations with AMS-02 at higher energy in the next two years might distinguish between these hypotheses.

Another source of excitement was the detection last year with the Fermi Gamma-ray Space Telescope of an excess of γ -rays near the Galactic centre, where dark matter should concentrate. A narrow spectral line at 130 GeV apparently associated with the

WHAT'S THE MATTER?

Dark-matter particles (D), such as weakly interacting massive particles and axions, can be spotted through their interactions with various types of standard-model particles (S) or with themselves. Experiments may detect them in four ways: directly; indirectly, by the particles such as photons they give off when they interact; in colliders; or through astrophysical observations.



LUX, Large Underground Xenon experiment; CoGeNT, Coherent Germanium Neutrino Technology Dark Matter Experiment; CTA, Cherenkov Telescope Array; HESS, High Energy Stereoscopic System; AMS, Alpha Magnetic Spectrometer; PAMELA, Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics; LHC, Large Hadron Collider. *Planned experiment or observations.



The LUX experiment detects photons produced when dark-matter particles interact with liquid xenon.

excess could indicate dark-matter-particle annihilations or decays. But a similar line from Earth's atmospheric limb implies that at least part of the signal must be instrumental in origin. A conclusive test may come in the next couple of years from the HESS (High Energy Stereoscopic System) γ -ray telescope in Namibia, which is observing the inner Galaxy in the 100 GeV to 1 TeV energy range.

The null results from LUX, the LHC and many other experiments are narrowing the range of possible particles that could explain dark matter. As claimed detections pop up only to disappear, physicists are becoming justifiably sceptical about every announcement of a discovery.

Some theorists have even started to wonder whether dark matter exists. Since the 1980s, a few have proposed modifying the theory of general relativity to do away with the need for dark matter. Such radical ideas are increasingly invoked to address another grave problem in astrophysics: the origin of the 'dark energy' that accelerates the expansion of the Universe. Most researchers think that we are far from needing new physical laws, especially because experimental avenues are still open. But unpleasant surprises are always possible.

There are two worst-case scenarios. First, dark matter may not comprise one type of particle — as many current searches assume — but many. Second, the particles might interact only gravitationally, and could be practically invisible to conventional detectors.

NEW DIRECTIONS

Existing experiments should run their course. But new approaches are needed to tease out dark-matter particles in the next decade.

A dark-matter modulation experiment,

such as DAMA/LIBRA or CoGeNT, in the Southern Hemisphere would gauge the extent of Earth's seasonal effects, which would be out of phase relative to the north.

Clumps or streams of dark matter moving through the Milky Way, distorting the rate at which particles hit detectors, should be visible as disturbances in the motions of the roughly one billion nearby stars that will be tracked by the European Space Agency's newly launched GAIA satellite during its five-year mission.

At the LHC and other next-generation accelerators, particle collisions with missing energy — drawn by an unknown particle — or other unexpected signatures could illuminate the dark sector.

We must also broaden directed searches and exploit astrophysical methods. First, we should look towards more massive particles, such as the SUSY particles. It will be difficult to detect heavy particles directly because there will be fewer of them. But γ -ray astronomy may help. The Cherenkov Telescope Array — an international project to build more than 100 ground-based telescopes to capture light flashes from γ -rays scattered by the atmosphere — should after 2015 open the window to 100-TeV energies. This energy coincides nicely with the highest limit on the WIMP mass expected from fundamental physics arguments. Such particles would generate TeV γ -rays when they annihilate or decay.

Second, broader categories of dark-matter particles should be sought. Like ordinary matter, dark matter could be complex, perhaps carrying a small charge, or having internal states akin to the electron levels of an atom. Changes in the Sun's oscillations as clouds of 'millicharged' particles scatter

off electrons in the solar plasma might be detectable through helioseismology. Gravitational lensing could measure the more-spherical dark haloes of distant galaxies, which are expected if dark-matter particles interact electromagnetically, albeit feebly.

Third is the axion. Predicted to explain an anomaly in quantum chromodynamics, the theory of the strong force, the electromagnetic signatures of axions have been long sought in the lab without success. String theory suggests types of ultralight axion that would be slightly more 'warm' than cold dark matter. Mixes of cold and warm dark matter, perhaps also including neutrinos¹⁰, might explain, for example, why there are fewer dwarf galaxies than cold-dark-matter scenarios predict.

Astrophysicists should look for unusual signals in old stars, such as neutron stars and white dwarfs. As stars orbit their galaxy, they accumulate WIMPs. Collected in the core of a neutron star, WIMPs might form a tiny black hole that could eventually devour the star, causing a violent explosion — an event that has yet to be seen. Helioseismology could also probe the effect of WIMPs on the Sun's temperature profile.

To refine theoretical and experimental strategies, particle physicists and astrophysicists need to communicate better. The number of dark-matter-candidate particles to be explored is limited, bounded at low masses by our failure to see anything and at high masses by the constraints of theory. A multidisciplinary approach to explore the 1–100 TeV mass–energy range should be the next frontier for the dark-matter community. ■

Mario Livio is an astrophysicist at the Space Telescope Science Institute in Baltimore, Maryland, USA. **Joe Silk** is professor of physics at the Institute of Astrophysics, Pierre and Marie Curie University, Paris, France. He is also in the Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland, USA; and is a senior fellow at the Beecroft Institute of Particle Astrophysics and Cosmology, University of Oxford, UK.

e-mail: mlivio@stsci.edu

1. LUX Collaboration. Preprint at <http://arxiv.org/abs/1310.8214> (2013).
2. Bertone, G., Hooper, D. & Silk, J. *Phys. Rep.* **405**, 279–390 (2005).
3. Bertone, G. *Nature* **468**, 389–393 (2010).
4. Goodman, M. W. & Witten, E. *Phys. Rev. D* **31**, 3059–3063 (1985).
5. Lockyer, N. *Nature* **504**, 367–368 (2013).
6. Bernabei, R. et al. *Eur. Phys. J. C* **73**, 2648 (2013).
7. Aalseth, C. E. et al. Preprint at <http://arxiv.org/abs/1401.3295> (2014).
8. Silk, J. & Srednicki, M. *Phys. Rev. Lett.* **53**, 624–627 (1984).
9. AMS Collaboration. *Phys. Rev. Lett.* **110**, 141102 (2013).
10. Viel, M., Becker, G. D., Bolton, J. S. & Haehnelt, M. *G. Phys. Rev. D* **88**, 043502 (2013).



Domestic goats and sheep can graze marginal lands, such as those in the Gobi Desert in Mongolia.

Steps to sustainable livestock

With improved breeding and cultivation, ruminant animals can yield food that is better for people and the planet, say **Mark C. Eislser, Michael R. F. Lee** and colleagues.

The need for efficient food production has never been greater. One in seven humans is undernourished¹. Urbanization and biofuel production are reducing land availability, and climate change, lack of water and soil degradation are decreasing harvests. Over the past decade, cereal yields per hectare have fallen in one-quarter of countries. Meanwhile, developing nations and the growing world population are demanding more animal protein.

The increasing consumption of animal protein is generally considered at odds with Earth's ability to feed its people. The 1 billion tonnes of wheat, barley, oats, rye, maize (corn), sorghum and millet poured annually into livestock troughs could feed some 3.5 billion humans. But such reasoning discounts the health benefits of eating modest amounts of meat and the fact that foraging animals can consume foods that humans cannot eat.

Crop and livestock farming complement each other². Half the world's food comes from farms that raise both. Animals pull ploughs and carts, and their manure fertilizes crops, which supply post-harvest residues to livestock. But efforts to maximize yields of milk and meat can disrupt finely balanced systems. The quest for 'intensification' in livestock

farming has thundered ahead with little regard for sustainability and overall efficiency (the net amount of food produced in terms of inputs such as land and water). With animal protein set to remain part of the food supply, we must pursue sustainable intensification and figure out how to keep livestock in ways that work best for individuals, communities and the planet.

Almost all of the world's milk and much of its meat come from ruminant (cud-chewing) animals — mostly cows, goats and sheep, but also buffalo, camels, llamas, reindeer and yaks. Here we highlight eight strategies to cut the environmental and economic costs of keeping these animals while boosting net gains for the quantity and quality of the food they produce.

Feed animals less human food. Around 70% of the grains used by developed countries are fed to animals. Livestock consume an estimated one-third or more of the world's cereal grain, with 40% of such feed going to ruminants, mainly cattle¹.

Some of this is avoidable. Ruminants graze pastures and can eat hay, silage and high-fibre crop residues that are unsuitable for human consumption. Unlike pigs, poultry

and humans, ruminants have a series of forestomachs leading to the true stomach. In the forestomachs, the largest of which is the rumen, microbes break down fibrous plant material into usable calories and also provide high-quality microbial protein. Ruminants can graze in marginal areas, such as mountainsides or low-lying wet grasslands. This helps to reserve agricultural fields for growing human food.

Even where large quantities of cereals are consumed by ruminants, up to 60% of their diet comes from high-fibre feed that humans cannot digest. In the European Union, more than 95% of milk comes from animals fed on grass, hay and silage, supplemented with cereals. Cattle in New Zealand's exemplary dairy industry obtain 90% of their overall nutrition by grazing pasture³. China's growing dairy industry initially relied on imported grain and high-quality fibre from the Americas. Ongoing research is showing how best to use local crop residues, such as rice straw.

Raise regionally appropriate animals. The lure of high productivity has led to ill-advised schemes to import livestock to places where they are genetically unsuited. Kerala, a state in southern India, is home to the smallest breed

COLIN MONTEATH, HEDGEHOG HOUSE/MINDEN PICTURES/FLPA

of cattle in the world. Vechur cows stand at about 90 centimetres tall and make only around 3 litres of milk per day — a dribble compared to the 30 litres per day produced on average by Holsteins, the black-and-white dairy cows of Europe and North America.

Donors, governments and charities aiming to feed whole communities, and to provide income for poor farmers, have imported Holstein breeding stock and semen to Africa and Asia, with progeny now numbering in the millions. But the animals often disappoint. Bred for centuries for maximum milk production in temperate climates, these cows were not selected for fertility or hardiness. They lack resistance to heat, humidity, tropical diseases and parasites, and so must be kept in stalls away from ticks and other disease vectors. Rather than allowing the animals out to pasture, farmers in tropical areas must cut and carry fodder to the animals or purchase expensive, often imported, feed. Even then, the cows produce less than one-third of yields seen in temperate climates and controlled environments. For poor families, a smaller native cow is a better bet than a larger animal that costs more to keep alive and healthy.

Similarly, breeds of cattle usually farmed in the humid tropics of West Africa have developed resistance to the debilitating disease trypanosomiasis over several thousand years of exposure to the tsetse fly that carries it. In the hope of greater profits and wealth, farmers often replace these animals with larger European cattle, or with zebu breeds from areas north of the tsetse belts. The zebu breeds are less resistant to trypanosomiasis, and European cattle have no resistance. The expense of drugs to combat the disease

often outweighs the increase in income.

More can and must be done to encourage farmers to realize the advantages of livestock adapted to local areas. Cutting-edge genomics could guide selective breeding to boost production of animals that are already adapted to their climates and resistant to local diseases.

Keep animals healthy. Sick animals can make people sick. In low- and middle-income nations, 13 livestock-related zoonoses (diseases that can infect humans and animals) cause 2.4 billion cases of human illness and 2.2 million deaths each year⁴. Yet human and livestock disease are generally treated as separate problems. Animal management should include measures to contain transmissible diseases, for example, by improving hygiene, quarantining new arrivals on farms and establishing coordinated, sustained surveillance for diseases that cross the boundaries of species or countries.

Mismanagement and poor welfare render animals particularly susceptible to parasites and disease. Many young animals die of disease before they can lactate, reach slaughter weight or reproduce. This lowers yields, increases environmental impacts and decreases farmers' ability to select the best breeding stock. With education and some financial aid, farmers could improve husbandry, and more animals would survive to become productive.

Keeping animals at high densities spreads infectious diseases far and fast. The foot-and-mouth virus costs upwards of US\$5 billion each year in vaccinations and lost production worldwide. A UK epidemic in 2001 resulted in the slaughter of 6 million animals. Bovine

tuberculosis has cost UK taxpayers alone £500 million (US\$830 million) over the past decade — an amount projected to double in the next ten years. Market disruptions and losses are felt across industries including agriculture, transportation and tourism.

European Union law holds farmers responsible for human health and food-safety issues following the slaughter of their animals. Growing awareness of problems such as antibiotic resistance has led to approaches that rely less on anti-infective drugs and more on management practices, such as reducing

overcrowding. Simple decision-support tools are emerging to help farmers to treat affected individuals rather than entire herds, and to keep animals away from risky pastures or other sources of infection⁵. Gathering local evidence can confirm the benefits of such strategies and encourage farmers to adopt them.

Adopt smart supplements. The productivity of ruminant animals can often be boosted with supplements, some of which encourage microbes in the rumen to grow quickly and to provide better nutrition. In India, a water fern (*Azolla caroliniana*) cultivated in local ponds provides extra protein to cattle and goats fed on protein-deficient elephant grass (*Pennisetum purpureum*).

Other plant extracts can alter the rumen microbial population to use nitrogen and energy more efficiently. This means producing more meat and milk with proportionally less by-product greenhouse gas and ammonia. An enzyme in red clover (*Trifolium pratense*), widely grown in temperate countries, increases ruminants' ability to utilize dietary protein⁶. In field trials, dairy cows with more clover in their diets ate more feed and produced more milk. In Australia, sheep nibble on the deep-rooted perennial tar bush (*Eremophila glabra*) during dry autumns when most other pasture plants offer poor food value. Tar bush combats gastrointestinal nematodes and acidosis, and reduces emissions of methane, a greenhouse gas 25 times more potent than carbon dioxide⁷.

Governments and policy-makers should support research efforts to identify the most beneficial microbes and most limiting nutrients, as well as low-cost ways to deliver them.

Eat quality not quantity. Annual meat consumption in India is just 3.2 kilograms per person, compared with 125 kg per person in the United States in 2007, much of it from heavily processed foods, such as burgers, sausages and ready meals. The focus should be on eating less, better quality meat. In rich countries, the high quantity and low quality of processed animal products consumed

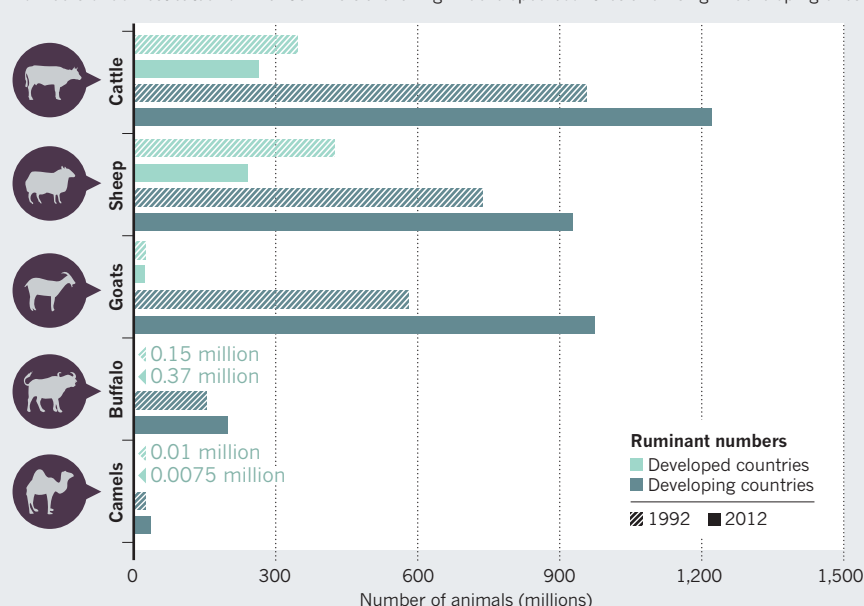
"The focus should be on eating less, better quality meat."

overcrowding. Simple decision-support tools are emerging to help farmers to treat affected individuals rather than entire herds, and to keep

SOURCE: FAO

CUD CHEWERS

Numbers of domesticated ruminant animals are falling in developed countries and rising in developing ones.





LEFT: TRAVEL INDIA/ALAMY; RIGHT: A. T. WILLET/ALAMY

Unlike Vechur cows (left), Holstein cattle (right) have little resistance to heat, humidity and tropical diseases, and are most productive in controlled environments.

contributes to ill health, with higher rates of cancer and coronary heart disease. For the world's poor people, however, there are clear nutritional advantages to consuming small amounts of high-quality animal foods, which are rich in protein, essential amino acids, iron and various essential micronutrients that improve chances for normal physical and cognitive development⁸.

The public-health goal, therefore, should be to balance nutrition across the world, with a target of weekly average consumption of red meat of no more than 300 grams. Trends are in the right direction; numbers of ruminants in the developed world have fallen over the last two decades (see 'Cud chewers').

Tailor practices to local culture. Close to one billion of the world's poorest people rely on livestock for their livelihood. Traditional animal husbandry supplies more than just food⁹. Keeping animals provides wealth, status and even dowry payments. When families encounter large expenses, such as a hospital bill or a wedding, they can sell an animal or two to cover the cost. Many of these benefits are disrupted when conventional grazing and mixed-farming practices are replaced with industrial systems that maximize short-term production. Policies to encourage humane, efficient management should consider cultural as well as natural factors. For instance, in traditional herding communities in the Horn of Africa, philanthropic efforts to support the cattle trade have led to larger herds for wealthier individuals, with little evidence that they have benefited poorer pastoralists.

Track costs and benefits. Livestock are widely considered to be unsustainable. The livestock sector accounts for 14.5% of human-induced greenhouse-gas emissions, exceeding that from transportation.

However, if other factors are considered, the picture becomes more favourable.

Sustainably managed grazing can increase biodiversity, maintain ecosystem services and improve carbon capture by plants and soil¹⁰. A cow produces up to 70 kg of manure per day, providing enough fertilizer in a year for one hectare of wheat, equivalent to 128 kg of synthetic nitrogen that might otherwise derive from fossil fuels. Mechanized arable agriculture and food processing themselves produce greenhouse gases, and costs of switching are exacerbated if meat's nutritional advantages are considered¹⁰. Farm animals also provide hides, wool, traction and biogas, a fuel produced from manure.

Calculating how this balances out is hard, but essential. Life-cycle assessment data should be used to tune livestock policies to socioeconomic and geographic environments.

Study best practice. To explore the multi-disciplinary strategies described here, we are building a global network of research farms. Three such 'farm platforms' are operational. Two focus on the use of naturally adapted livestock and native plants: the University of Western Australia Future Farm in Pingelly, which has Mediterranean biome conditions and where water conservation is crucial, and the Thiruvazhamkunnu Livestock Research Station in Kerala, India, which has humid tropical conditions and where grazing is strictly limited. At the third, Rothamsted Research North Wyke Farm Platform in Devon, UK, cattle and sheep graze in temperate grassland conditions on three hydrologically isolated, 22-hectare 'farmlets' to compare nutrient cycling and productivity under various pasture-management strategies. There are plans to establish further platforms in South America, North America and China.

There will be no one-size-fits-all solutions. Changing farming practices is difficult, but farm platforms can evaluate potential for increased profits and other benefits, act as

examples to follow, and provide information for policy-makers. We hope to identify better practices to optimize the use of livestock in different regions, using local resources, breeds and feedstuffs — and produce tangible evidence to convince local farmers. ■

Mark C. Eisler is professor of global farm animal health, **Michael R. F. Lee** is reader in sustainable livestock systems and **John F. Tarlton** is senior research fellow in matrix biology at the School of Veterinary Sciences, University of Bristol, UK.

Graeme B. Martin is professor of animal science at the UWA Institute of Agriculture, University of Western Australia, Perth, Australia.

John Beddington, Jennifer A. J. Dungait, Henry Greathead, Jianxin Liu, Stephen Mathew, Helen Miller, Tom Misselbrook, Phil Murray, Valil K. Vinod, Robert Van Saun, Michael Winter.

e-mail: mark.eisler@bristol.ac.uk

1. Food and Agriculture Organization of the United Nations *World Agriculture: Towards 2015/2030* (FAO, 2002).
2. Herrero, M. *et al. Science* **327**, 822–825 (2010).
3. Bocquier, F. & Gonzalez-Garcia, E. *Animal* **4**, 1258–1273 (2010).
4. Grace, D. *et al. Mapping of Poverty and Likely Zoonoses Hotspots* (International Livestock Research Institute, 2012).
5. Kenyon, F. *et al. Vet. Parasitol.* **164**, 3–11 (2009).
6. Lee, M. R. F., Tweed, J. K. S., Minchin, F. R. & Winters, A. L. *Anim. Feed Sci. Technol.* **149**, 250–264 (2009).
7. Bickell, S. L., Durmic, Z., Blache, D., Vercoe, P. E. & Martin, G. B. in *Updates on Ruminant Production and Medicine Proc. 26th World Buiatrics Congress, Santiago, Chile* (eds Wittwer, F. *et al.*) 317–325 (Andros Impresores, 2010).
8. Smith, J. *et al. Anim. Front.* **3**, 6–13 (2013).
9. Otte, J. *et al. Livestock Sector Development for Poverty Reduction: An Economic and Policy Perspective — Livestock's Many Virtues* (FAO, 2012).
10. Garnett, T. *Environ. Sci. Policy* **12**, 491–503 (2009).

Full author affiliations accompany this article online at go.nature.com/8d9vjx.

FICTION

The cryptic mind

Li Gong ponders a translation of a Chinese novel on the world of code-breaking.

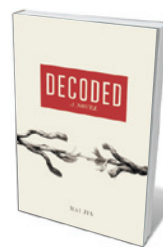


Bestselling novelist Mai Jia is alleged to have worked in cryptography.

An autistic boy is born out of wedlock, descended from a mathematical-genius grandmother who helped the Wright brothers to design their first planes. The boy is a maths prodigy himself, and eventually becomes the top code-breaker in the Chinese military — but is tormented by the psychological contradictions of cryptography. However implausible-sounding the story of Mai Jia's breakthrough 2002 novel *Jie Mi* (*Decoded*), this is a spy thriller grounded in subtle and difficult realities. Mai — one of China's pre-eminent writers — is alleged to have worked in cryptography himself. The battle in the reader's mind over whether this is florid drama or brutal realism is perhaps the primary and primal attraction of the novel, translated into English for the first time by Olivia Milburn and Christopher Payne.

Decoded is written in the style of a film script, with rapid cutting between scenes and the reader thoroughly gripped by the characters' tribulations. It is also necessarily dressed up with copious references to mathematics — names of mathematicians, theorems, and encoding and decoding methods — and even includes a few formulae. Yet its main theme is human psychology, and especially the tortured psyches of those who, like protagonist Rong Jinzhen, spend their lives trying either to hide information or to crack the protective puzzles of their enemies. Rong's old tutor, for instance, writes to his former protégé, condemning decryption and cipher construction as "fundamentally anti-scientific, anti-intellectual ... a poison that mankind has developed to destroy science and a conspiracy against the people that work with them".

The translation is good overall, but there are a few fatal errors. For example, Mai takes a swipe at Chinese higher education, pointing out that a major problem with Chinese academics is that they start out as scholars and end their careers as government officials; he hints that they regard this path as



Decoded: A Novel
MAI JIA (TRANSLATED
BY OLIVIA MILBURN
AND CHRISTOPHER
PAYNE)
Allen Lane/Farrar,
Straus & Giroux: 2014.

GRAYHAWK AGENCY

a most desirable and natural progression. As the translation has it, however, Chinese academics regard scholarship and officialdom as incompatible — the exact opposite of Mai's meaning.

Another lapse undermines a central tenet of the novel: the supposed accepted wisdom in cryptography that one person can only design, or crack, one good cipher at most. This is because cryptographers can be made vulnerable both by subconscious similarities in their own ciphers, and by the exposure of their specific code-breaking strengths when they crack other people's ciphers. Yet the English version translates this as it being impossible for one person to be both a cipher designer and a code-breaker. As a result, readers might not fully understand the tragedy of the story — that a man who successfully broke the enemy's sophisticated Code Purple goes mad at failing to break its successor.

With this and subsequent novels such as the 2003 *An Suan* (*In the Dark*, forthcoming from Penguin) selling in their millions and being adapted into high-profile television series, Mai has firmly established himself as the father of modern Chinese espionage thrillers. Critical acclaim for his work has included China's most prestigious book award, the Mao Dun Literature Prize.

So is Mai — the supreme storyteller of psychological warfare, intrigue and the human sufferings of alienation — a brave writer who unveils national security programmes and pinpoints untold human sacrifices in a secretive state? Or is he a commercial master who plays on the public's desire to

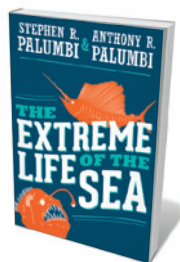
“Decoded is dressed up with references to mathematics, yet its main theme is human psychology.”

peer into a covered-up world in a country where media is officially controlled and pumps out content-free content every day? Is he exploiting the public's wish to believe that an outrageous, even over-dramatized story is the best vehicle through which to tell an untellable truth?

Even publishers had difficulty answering this question. As Mai relates in an appendix to a 2011 Chinese edition, he first submitted the manuscript of *Decoded* to two friendly publishing houses. Both editors quickly rejected his work. One thought the story too fake to attract interest; the other felt it was too real and sensitive, and could cause trouble for the publisher. Decoding Mai Jia's real intentions is perhaps the book's most perplexing challenge. ■

Li Gong is President of Asia Operations at Mozilla Corporation in Beijing.
e-mail: lgong@mozilla.com

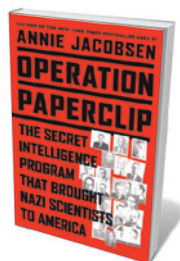
Books in brief



The Extreme Life of the Sea

Stephen R. Palumbi and Anthony R. Palumbi PRINCETON UNIVERSITY PRESS (2014)

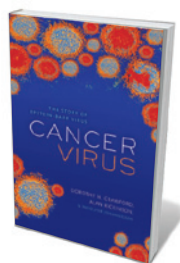
Marine biologist Stephen R. Palumbi and writer Anthony R. Palumbi survey an impressive catch of extreme oceanic species, from the oldest to the deepest-dwelling. They are inspired guides, weaving evolutionary and geological backstories into accounts of wonders such as the exquisite architecture of sharks' teeth. And they pull no punches in depicting potential futures with devastated oceans dominated by “bacteria, jellyfish and tar-like algae”. A brilliant use of the rich store of research into Earth's largest habitat.



Operation Paperclip: The Secret Intelligence Program that Brought Nazi Scientists to America

Annie Jacobsen LITTLE, BROWN (2014)

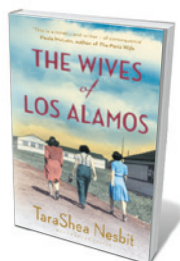
An ethical murk can hang over science in service to the state, as Annie Jacobsen reveals in this chilling history. Operation Paperclip united Nazi science and US cold-war interests, bringing 1,600 German technologists to the United States after the Second World War to work on intelligence and weapons research, despite protests from the likes of Einstein. Jacobsen focuses on 21 scientists, including rocket pioneer and Third Reich insider Wernher von Braun, and reveals disturbing evidence of Paperclip's legacy in US science and medicine.



Cancer Virus: The story of Epstein-Barr Virus

Dorothy H. Crawford, Alan Rickinson and Ingólfur Johannessen OXFORD UNIVERSITY PRESS (2014)

Fifty years ago, the discovery of a virus that triggers cancer in humans rocked the medical world. In this pithy, pacy study, the Epstein-Barr virus is biographized by three scientists who work on it — Dorothy H. Crawford, Alan Rickinson and Ingólfur Johannessen. Starting with the discovery of Burkitt's lymphoma in the 1950s and the isolation of the virus by Michael Anthony Epstein and Yvonne Barr in 1964, they follow its trail to effects in China and the labs where it has proven research gold for molecular biologists and geneticists.



The Wives of Los Alamos: A Novel

TaraShea Nesbit BLOOMSBURY (2014)

Behind the men behind the Los Alamos nuclear-research programme were women whose story has been waiting to be told. TaraShea Nesbit has done it lyrically in this novel. Written in the collective voice of “the wives” — international, often highly educated women — this chronicle of the Manhattan Project's secret wartime base in New Mexico unfurls as they lived it, distorted by necessary lies. Their strange existence as housewives, “calculators” or lab technicians forms a vivid foreground to the building of the bombs that finished the Second World War, with Nesbit deftly capturing the claustrophobic surreality of it all.



Ha!: The Science of When We Laugh and Why

Scott Weems BASIC BOOKS (2014)

Mirth, points out cognitive neuroscientist Scott Weems, is still something of a conundrum — but one well worth cracking. His journey through the jovial looks in turn at what it is, what it is for and why we should cultivate it. We encounter British psychologist Richard Wiseman's LaughLab and its findings (Americans laugh at insults; Europeans savour the absurd), the power of conflict and messy thinking, the speculation that atheists are funnier, the beneficial impact of laughter on pain tolerance, and more.

PALAEOLOGY

Migrant nation

Ewen Callaway revels in an exhibition chronicling the ebb and flow of early humanity over the British Isles.

Despite what some politicians say in these inward-looking times, Britain is and always has been a nation of immigrants. Even the palest Brits descend from Middle Eastern stock or other waves of foreigners, beginning with the arrival of agriculture. *Britain: One Million Years of the Human Story* — an enlightening and engaging exhibition at the Natural History Museum (NHM) in London — lays out the evidence for an immensely long history of migration, ebbing and flowing over the millennia.

The exhibition is the fruit of a 13-year research project led by the NHM, the Ancient Human Occupation of Britain. It shows that Britain hosted ten distinct occupations of various hominin species (described as humans throughout the exhibition) between around 1 million and 12,000 years ago. “Nine of them died out, and we are the tenth,” says Chris Stringer, research leader in human origins at the NHM.

For many, the show’s biggest appeal will be the geography, not the genealogy. Those 900,000-year-old hand-axes once used to butcher big game? They were found near a wide sandy beach in Norfolk, popular with dog-walkers. That 200,000-year-old hippopotamus skeleton? Pigeons now pace Trafalgar Square in the heart of London, where it was dug up. For an exhibition about the human past, there are surprisingly few human bones on display. This is not a curatorial oversight, but a consequence of the dearth of human fossils in Britain. The show may be better for it. Because it is forced to rely on evidence such as stone tools and marked animal bones, it demonstrates how researchers make inferences about the people who left such artefacts behind.

Take the first humans to inhabit Britain, more than 900,000 years ago. An intriguing find revealed just last month shows that they may have belonged to the species *Homo antecessor*, first discovered at Atapuerca in northern Spain. A team including Stringer found footprints embedded in stone near Happisburgh on the Norfolk coast, and speculated that the appearance and the gait of the humans who made

them would have been similar to those of *H. antecessor*. A short documentary film and photo display of the footprints — now eroded away by the sea — are shown.

The next humans to call Britain home may have been *Homo heidelbergensis*, known from fossils in Germany and elsewhere. This exhibit points to large flint hand-axes found alongside the butchered remains of a rhinoceros near Boxgrove, West Sussex. A film re-enacting the grisly slaughter threatens to overshadow a fascinating exhibit nearby: electron micrographs of the scored rhino

Britain: One Million Years of the Human Story
Natural History Museum, London.
Until 28 September 2014.

bones, which helped to show that the damage was caused by human butchery rather than environmental weathering.

Neanderthals, the big-bodied hunters who lived in Europe and western Asia, came at least twice to the British Isles, which were intermittently connected to continental Europe. The Neanderthals’ first stretch in Britain, 400,000–200,000 years ago, was interrupted by a harsh cold snap that made the area uninhabitable for humans. The rise of waters with rapid warming then turned Britain into an island, preventing continental Neanderthals from returning until alternative routes opened up. Their second visit, starting less than 100,000 years ago, may have been interrupted by anatomically modern humans who arrived from Africa some 60,000 years later.

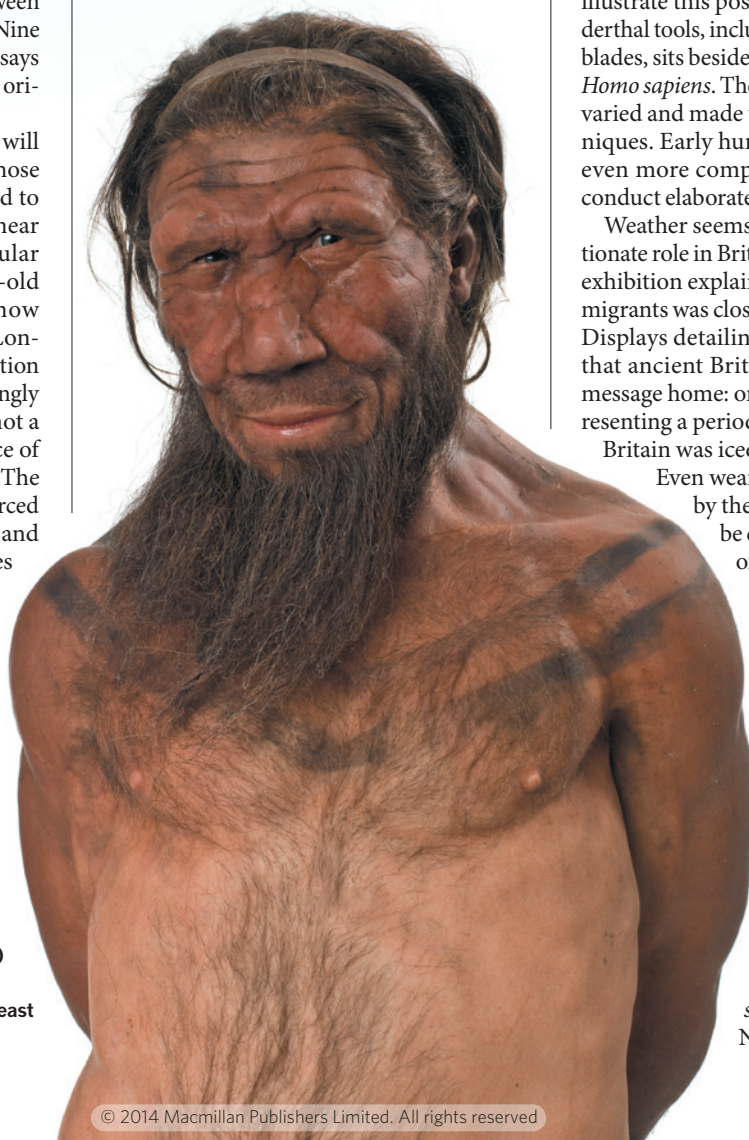
Many details of the relationship between Neanderthals and modern humans stand on scientifically thin ice, but the exhibition mostly avoids just-so stories (Did they battle? Did they breed?). Instead it presents a nuanced hypothesis, influenced by Stringer’s thinking, that Neanderthals could not or did not adapt to changing climate and competition for resources, and slowly died out. To illustrate this possibility, an array of Neanderthal tools, including hand-axes and stone blades, sits beside implements fashioned by *Homo sapiens*. These tend to be lighter, more varied and made using more efficient techniques. Early humans went on to develop even more complex tools and art, and to conduct elaborate burials.

Weather seems to have had a disproportionate role in British life from the start. The exhibition explains how the rise and fall of migrants was closely tied to climate change. Displays detailing the plants and animals that ancient Brits encountered drive this message home: one is an empty room, representing a period 450,000 years ago, when Britain was iced over and uninhabitable.

Even weary museum-goers enticed by the empty benches will soon be chased off by a soundtrack of howling winds.

The discoveries unearthed by the Ancient Human Occupation of Britain project have twice pushed back the date of Britain’s earliest human occupation. “With further work there could be even older evidence of humans in Britain,” Stringer says. Revel in it: yet more immigrants. ■

Ewen Callaway is a senior news reporter for *Nature* in London.



Neanderthals (reconstruction) migrated to the British Isles at least twice.



Santiago Calatrava's auditorium in Tenerife, Spain: tinted concrete flecked with white ceramic.

Q&A Andrea Hamilton

Concrete conservator

Materials scientist and engineer Andrea Hamilton at the University of Strathclyde in Glasgow, UK, uses chemistry research to conserve the structural and aesthetic integrity of concrete. She talks about dams, nuclear-waste storage and an artwork that explores weathering.

NICK HIGGINS



How did you get interested in materials engineering?

I have a very emotional response to architecture. Cement and concrete can be seen as synonymous

with poverty and degradation, even war. But when I started to study it, I gained a greater aesthetic appreciation.

What are some examples of beautiful concrete structures?

The Catalan architect Antoni Gaudí used cement and concrete in novel ways. Rather than modular cookie-cutter structures, you see curvaceous, colourful forms because precast, reinforced concrete allowed him to push the boundaries of structural engineering. Stylistically, my favourite is an auditorium by Santiago Calatrava in Tenerife, Spain. Made from steel-reinforced concrete, it looks like a huge wave, curving downward into a narrow point, and its 3,500-tonne weight is supported by just five points of contact. The surface is reflective; Calatrava tinted the concrete using

titanium dioxide and decorated the surface with fine pieces of white ceramic.

Why is concrete so hard to conserve?

Cement is chemically complex — more so than any other building material. Portland cement, the most common type, is made from clay and limestone, heated to around 1,450°C. This forms clinker, containing at least 70% calcium silicate minerals, which is ground and mixed with a small amount of gypsum to create cement powder. Mixed with water and aggregate, that makes concrete. There are lots of variables involved, particularly in mix design. I have worked on the conservation of cement-rendered brick buildings at the East Fortune airfield in Scotland. Built during the Second World War, they were not intended to last. Frost damage has made the cement render peel.

How has ancient concrete survived?

The Romans must have experimented. They used lime mixed with volcanic ash, producing a relatively aluminium-rich system. Recent research has shown that Al-tobermorite [a calcium silicate hydrate mineral] formed in ancient harbour concrete such as a breakwater in Pozzuoli Bay near Naples. This

enhanced the concrete's durability, even in sea water. The Romans also added animal hair as fibre reinforcements, and blood for frost resistance — both excellent innovations.

How can we make concrete more environmentally friendly?

Every tonne of Portland cement produced generates about 1 tonne of carbon dioxide, half of it from the chemical reaction when you heat limestone. You can't avoid carbon dioxide coming off when you heat limestone, but you can control how much Portland cement goes into a mix. To me, the most promising research on the ecological side focuses on incorporating more industrial by-products. Up to 80% of the cement can be replaced with fly ash or slag, although it takes longer to set.

How can science aid concrete's longevity?

Early research explored the bulk mechanical properties of concrete. Now research tends to focus on microstructures and crystallography to tailor clinker reactivity. Controlling permeability is also essential, as one of the main problems associated with concrete and cement is water ingress. With colleagues at Strathclyde and the Nuclear Decommissioning Authority, I'm working on structures to house nuclear waste. One member of the project is developing smart sensors that can be embedded in concrete to detect properties such as water and chloride content, and temperature, in real time. My goal is to use this data to develop and feed a real reactive transport model of concrete deterioration.

Do concrete-rich countries such as the United States present particular challenges?

The dramatic feature of Frank Lloyd Wright's iconic 1930s house Fallingwater is reinforced concrete 'trays' cantilevered out over a waterfall; in the late 1980s the cantilevers had to be supported from within because they were sagging. Of US infrastructure, dams are an issue: their average age is 52 years; it could cost more than US\$50 billion to rehabilitate them all.

Are you working on any art projects?

I'm involved in *Salt Licks*, a fascinating project due to be built this summer on the Lincolnshire coast. It will be a 6.6-metre-high reinforced concrete construction, clad with salt blocks. The idea is that the salt will dissolve, allowing a dynamic shape to form. This represents a fantastically severe test of how concrete weathers in contact with salt. Time-lapse digital cameras will monitor it. We will see the immediate impact of driving rainfall, in patterns that have never been visible before. My interest is in informing new art. Once you have a bag of cement, aggregate and water plus guidelines, you can create almost anything, anywhere. ■

INTERVIEW BY ALEXANDRA WITZE

Correspondence

Synthetic ivory fails to stop illegal trade

Synthetic ivory can now be crafted to the same diagnostic standards as genuine ivory (see M. E. Sims *et al. Ethnobiol. Lett.* **2**, 40–44; 2011), and its price in China is only about 14% of that of real ivory. First manufactured in 1865 to save elephants' tusks from being turned into billiard balls (US patent 50359), synthetic ivory is not proving to be the panacea hoped for by conservationists and Chinese enforcement agencies.

Because synthetic and authentic ivory are so similar, unscrupulous traders caught smuggling illegal ivory can claim that it is synthetic; they can also pass off synthetic ivory as genuine when they sell it. The situation may be aggravated by legitimate traders, because they are entitled to compensation if destructive sampling is carried out to conclusively distinguish real from synthetic ivory (H. G. M. Edwards and D. W. Farwell *Spectrochim. Acta A* **51**, 2073–2081; 1995).

One of us (Z.-M. Z.), as an enforcement officer for the Yunnan Public Security Bureau for Forests in China, has investigated 57 cases of suspected illegal ivory trading since 2011. Of these, 27 attempted to disguise samples of genuine ivory by mixing them with fake ivory, and only 513 of 1,714 items actually proved to be synthetic. These litigants all voluntarily submitted their products to destructive sampling, saying that they were deceived by their own suppliers.

The illegal ivory trade does not seem to be diminishing: in November 2013 alone, for example, customs agents in Xiamen, China, seized around 12,000 kilograms of imported ivory, and three Chinese citizens were arrested for smuggling 1,800 kg of ivory from Tanzania. **Zhao-Min Zhou*** *Yunnan Public Security Bureau for*

Forests, Kunming, China.

zhouzm81@gmail.com

**On behalf of 4 co-signatories; see go.nature.com/prbm3o for full list.*

GDP: signposting a false choice

Robert Costanza and colleagues suggest that it is time to abandon gross domestic product (GDP) as a measure of national success (*Nature* **505**, 283–285; 2014). However, I feel that your choice of an illustration of a signpost to make this point conveys the wrong message — namely, that we must decide between economic growth and sustainable progress.

The authors' point is that GDP growth does not equate to economic growth: it is a measure of currency transactions but not of net value creation. Economic growth, properly considered, is an increase in net value produced and a rise in capital assets that can be used for value production.

Modern scientific thinking classes natural ecosystems as capital assets, so we should include the protection of ecosystems as value production and the destruction of ecosystems as value depletion. Likewise, improvements in social systems and general social well-being create value.

The genuine progress indicator (GPI) incorporates economic growth as defined in these terms, so there is no dichotomy: the signposts should be pointing in the same direction.

Bryan Long *Palo Alto, California, USA.*

bryan@ecosystemics.com

GDP: no one metric can rule them all

Robert Costanza and colleagues call for the development of an alternative metric that includes a broader set of information than gross domestic product (GDP) and better reflects human well-being (*Nature* **505**, 283–285; 2014). Improved metrics are

needed, but we believe that more understanding is gained from balancing many related metrics.

The authors tell a common story of too much reliance on a single metric and too little attention to its shortcomings as it shapes our world view and influences behaviour. For instance, current US political discussion is heavily focused on a 'jobs' metric, which also conveys an incomplete, and often misleading, message. As with GDP, this metric does not distinguish between activities that help or hurt well-being, and it does not offer much insight into social health and sustainability.

Individually, metrics can mislead and misdirect; in a set of complementary metrics, they can provide invaluable knowledge. Composite metrics are useful, but embody weighted components that impose and conceal trade-offs that should be understood and discussed.

Ironically, as we increasingly understand the complexities of how the world works, policy-makers and the public want to distil this complexity into a single metric. Improving a metric is a step forwards; improving our use of metrics would be a leap forwards.

Gregg Marland, Todd L. Cherry, Howard S. Neufeld *Appalachian State University, Boone, North Carolina, USA.*
marlandg@appstate.edu

Addiction: many factors contribute

Your assertion that drug addiction is a brain disease (*Nature* **506**, 5; 2014) is a contentious oversimplification — akin to always blaming a car crash on the car alone.

By this analogy, addiction is the car crash and the brain is the vehicle: the person driving it contributes a psychological component, the others on the road represent social factors, and road conditions correspond to the environment. To promote any

one of these as the prime cause of addiction, as in the brain-disease model, overlooks the importance of the other components.

There is a risk that such a reductionist biological view could hinder the development of solutions for addiction. It could also undermine a considered response to the issue of politically driven bans on the use of animals in addiction research — itself a topic that has important social, psychological and environmental contributors.

John A. Cunningham *Australian National University, Canberra, Australia.*
john.cunningham@anu.edu.au

Addiction: not just brain malfunction

Irrespective of the animal-rights issues you discuss, we disagree with your one-dimensional view that addiction is a disease, and with your claim that this view is not particularly controversial among scientists (*Nature* **506**, 5; 2014). Neuroscience has been widely documented as just one of many important influences in drug addiction.

Substance abuse cannot be divorced from its social, psychological, cultural, political, legal and environmental contexts: it is not simply a consequence of brain malfunction. Such a myopic perspective undermines the enormous impact people's circumstances and choices have on addictive behaviour. It trivializes the thoughts, emotions and behaviours of current and former addicts. It also belittles the significant effect that public-health campaigns and legislation exert on curbing substance abuse.

Addiction is too complex to be fought on a medical-research front alone. A variety of approaches based on diverse levels of analysis is required.

Derek Heim* *Edge Hill University, Ormskirk, UK.*
heimd@edgehill.ac.uk
**On behalf of 94 co-signatories; see go.nature.com/ehujzh for full list.*

EARTH SCIENCE

Missing link in mantle dynamics

The discovery of crystallographic imperfections known as disclinations in the most profuse mineral in Earth's upper mantle has the potential to solve a problem that has vexed mineral physicists for decades. [SEE ARTICLE P.51](#)

GREG HIRTH

The viscosity of Earth's upper mantle controls a wide range of processes, from the attenuation of seismic waves and the rate of surface deformation after earthquakes to the slow, global-scale flow that is associated with mantle convection and the dynamics of tectonic plates. This viscosity is logically interpreted to be dominated by the physical properties of olivine, the most abundant mineral in Earth's upper mantle, as well as in those of the other terrestrial planets (Mars, Venus and Mercury) and the Moon. In this issue, Cordier *et al.*¹ (page 51) report how new techniques to analyse the microstructure of grain boundaries in olivine allowed them to discover crystal defects called disclinations in this mineral. This observation is probably a first for geological materials, and has ramifications for our understanding of the processes that control mantle dynamics.

More than 50 years of effort have gone into the experimental and theoretical characterization of the solid-state flow properties of olivine. Motivated by the principles of materials science, a sizeable fraction of these studies has concentrated on understanding the relationships between the mobility of crystal defects, primarily dislocations, and the creep behaviour (deformation that occurs under continued stress) of both olivine single crystals and olivine aggregates, which are known as peridotites.

Owing to the scales of deformation in the Earth, the results of experiments on olivine must be extrapolated by many orders of magnitude in spatial scale, timescale or stress. Nonetheless, excellent agreement exists between predictions based on such extrapolation and independent geophysical observations. Furthermore, the details of microstructures in mantle samples collected from mountain belts and xenoliths (pieces of rock from great depth that are brought to the surface in volcanic eruptions) show many striking similarities to those observed in experimental samples, providing a first-order validation for the extrapolation. However, the application of these relationships to mantle dynamics remains compromised because a fundamental principle for grain-scale deformation of aggregates, the Von Mises criterion, has been hard to

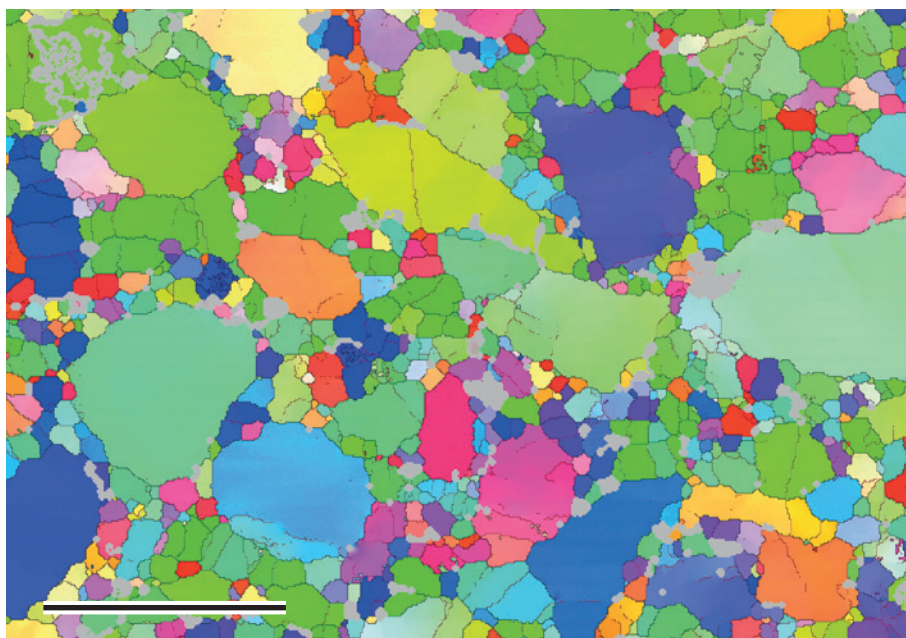


Figure 1 | Imaging polycrystalline olivine. Cordier *et al.*¹ have analysed images produced using electron backscatter diffraction to identify grain-boundary disclinations in olivine aggregates. Shown here is a cross-section of a sample deformed at high stress. Different colours denote different crystallographic orientations of the crystals that make up the aggregate. Scale bar, 20 μm .

reconcile with these laboratory experiments.

The Von Mises criterion states that five independent slip systems — combinations of crystal planes and directions in which defects move — are required to allow homogeneous grain-scale deformation of polycrystalline aggregates; this requirement can be relaxed to four slip systems for inhomogeneous grain-scale deformation². However, extensive work on olivine single crystals suggests that only three independent slip systems exist for reasonable stress states in both laboratory and natural conditions. This has led to suggestions that deformation accommodated by the diffusion of mineral components through a process called dislocation climb, or through grain-boundary sliding^{3,4}, might account for the 'missing' grain-scale deformation required by the Von Mises criterion. These hypotheses are challenged by the extremely sluggish diffusion kinetics observed for olivine, and the limited range of grain size where evidence for grain-boundary sliding is observed. Now Cordier *et al.* suggest that the motion

of grain-boundary disclinations could be a solution to the problem.

As noted by Cordier and colleagues, the idea of disclinations was first described⁵ more than 100 years ago. However, awareness of their significance has been growing in recent years, owing to advances in microscopy⁶ and continued development of theory for the defect structure of grain boundaries⁷. Disclinations (and dislocations) can be visualized as linear defects related to the distortion of a cylinder; for example, a twist disclination is defined by a rotation on an axis perpendicular to a cut that runs along the longitudinal axis of the cylinder (see Fig. 1 of the paper¹). Stress concentrations around these linear defects allow intracrystalline deformation at stresses much below the theoretical stress required to break a plane of atomic bonds in the crystal. In the present study, the authors took advantage of the high-resolution crystallographic technique of electron backscatter diffraction to identify grain-boundary disclinations in olivine (Fig. 1).

SYLVIE DEMOUCY

Many questions remain about the efficacy of the disclination hypothesis for resolving the Von Mises problem in peridotites. For example, how does their density and mobility evolve during deformation and with changes in thermodynamic conditions, and how are the dynamics of disclinations influenced by the anisotropic (direction-dependent) elastic properties of olivine? How are grain-boundary disclinations involved with grain-boundary sliding? Can the inclusion of disclinations in polycrystalline-deformation models help to resolve questions regarding the evolution of lattice-preferred orientations

in olivine aggregates deformed to high deformation? The latter problem is crucial for the interpretation of the seismic structure of Earth's mantle, which in turn is currently our best observational technique to explore convective motions inside Earth. Research on these topics promises to provide insights into the viscosity of the terrestrial planets and into the interpretation of a broad range of geophysical observations. ■

Greg Hirth is in the Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA.

e-mail: greg_hirth@brown.edu

1. Cordier, P. *et al.* *Nature* **507**, 51–56 (2014).
2. Hutchinson, J. W. *Met. Trans. A* **8**, 1465–1469 (1977).
3. Hirth, G. & Kohlstedt, D. L. *J. Geophys. Res.* **100**, 15441–15449 (1995).
4. Hansen, L. N., Zimmerman, M. E. & Kohlstedt, D. L. *J. Geophys. Res.* **116**, B08201 (2011).
5. Volterra, V. *Ann. Sci. Ecole Norm. Sup.* **24**, 401–517 (1907).
6. Murayama, M., Howe, J. M., Hidaka, H. & Takaki, S. *Science* **295**, 2433–2435 (2002).
7. Hirth, J. P., Pond, R. C. & Lothe, J. *Acta Mater.* **54**, 4237–4245 (2006).

This article was published online on 26 February 2014.

BIOLOGICAL TECHNIQUES

Wrapped around the heart

An elastic membrane cast around a three-dimensional printed model of a specific heart allows diverse aspects of cardiac function to be monitored and modified, and paves the way to new diagnostic and therapeutic approaches.

COLLEEN E. CLANCY & YANG K. XIANG

A long-sought goal has been to observe, diagnose and treat dangerous disturbances to cardiac rhythm, an unfortunately common situation that occurs when electrical impulses in the heart become disordered. The major roadblock in the quest to understand and treat abnormal cardiac rhythms (arrhythmias) is that there is no straightforward approach for detailed inspection of the underlying electrical, metabolic and mechanical changes and the interactions between them. Writing in *Nature Communications*, Xu *et al.*¹ have begun to address this deficit by developing three-dimensional membranes that uniquely conform to the heart surface, providing a platform for flexible arrays of multifunctional sensors.

The devices, termed three-dimensional multifunctional integumentary membranes (3D-MIMs), offer several key features. They are inherently elastic and mechanically stable, and they provide a uniform, non-invasive interface to all points on the heart, allowing high-resolution measurement or high-density stimulation of cardiac function (Fig. 1). Furthermore, the components of the attached sensors can be distinctively configured to allow a variety of functions, including electrical, thermal and optical stimulation, or measurement of pH, temperature, voltage or strain.

The 3D-MIMs are made by casting a thin layer of silicone elastomer (a silicone-containing, rubber-like material) over a 3D-printed reconstruction of the heart of interest, which

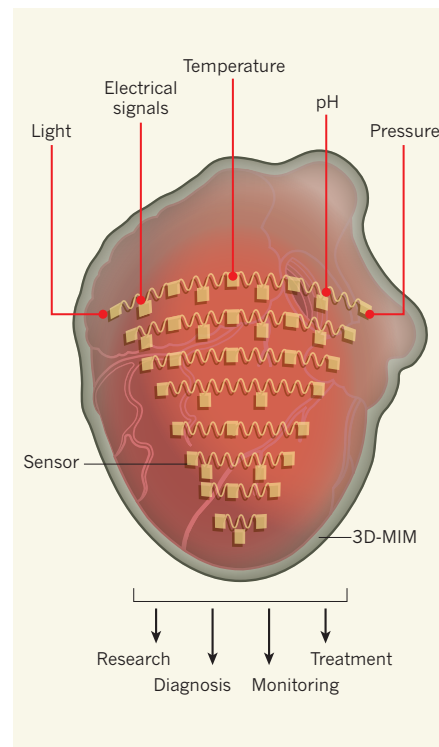


Figure 1 | Multifunctional monitoring. Xu *et al.*¹ have developed three-dimensional elastic membranes (3D-MIMs) that conform to the surface of a specific heart, creating a platform for arrays of multifunctional sensors and electronic and optoelectronic components that can stimulate and measure the pH, temperature, voltage or strain in the heart. These devices could be used in research applications and in the clinic for diagnosis, monitoring and treatment.

is generated using scanning techniques involving optical segmentation, magnetic resonance imaging or computed tomography. This process produces an artificial envelope that uniquely conforms to a specific heart. The manufacturing technique can be applied to different species and, importantly, to individual patients.

One of the immediate research applications of 3D-MIMs will be to perform multiple simultaneous measurements of a working (beating) heart during normal or abnormal cardiac cycles. Such measurements include temperature, tension and pH, any or all of which can be recorded simultaneously from electrical signals. Currently, optical mapping with voltage-sensitive dyes is the best available technique for measuring electrical function with high spatial resolution in the intact heart. However, one of the acknowledged shortcomings of this approach is that it relies on excitation–contraction uncoupling drugs that allow recording only in the motionless heart. 3D-MIMs overcome this limitation, allowing the first possibility of high-resolution measurements of excitable and metabolic states in the beating heart.

Different actuators for electrical, thermal and optical stimulation can also be embedded in 3D-MIMs. This would allow the potential integration of optogenetic tools² (involving light-responsive proteins) or genetically coded biosensors³ for monitoring electrical and biochemical signals, or to locally activate or deactivate genes or proteins in normal or diseased hearts. Furthermore, such measurements or interventions could be applied to hearts in the resting state or when they are subject to various stressors.

In the clinical realm, 3D-MIMs have almost limitless potential for the construction of patient-specific devices for diagnostic and therapeutic purposes. During open-heart surgery, 3D-MIMs with various combinations of components will be useful for diagnosing the region of an abnormality, such as arrhythmias, ischaemia or heart failure, which will help to determine appropriate surgical or drug treatment strategies. The membranes might even be used to deliver therapeutic interventions during

surgery, including localized tissue ablation, pacing or acute low-energy defibrillation⁴.

Although the immediate potential for using 3D-MIMs in such open-chest surgical scenarios is clear, chronic implantation of 3D-MIMs in patients could prove especially effective. Such implantation might eventually allow for post-surgery monitoring involving continuous measurements of local metabolic, excitable, ionic, contractile and/or thermal states of the heart in response to various insults, diseases or therapies. These diagnostic data might then be used to activate the device remotely to deliver targeted electrical therapy without the need for further surgery. The membranes might also be developed as a platform for low-energy control or defibrillation methods to regulate electrical turbulence

in the heart⁴. In the same vein, 3D-MIMs could conceivably be used for localized and targeted delivery of stem cells, viral vectors or drugs.

There is also the future possibility of creating an *in vivo* optical-mapping system using externally applied dyes or internal fluorescent indicators. This application would require the development of diverse light sources that could be integrated into the membranes and provide enough power to illuminate both dyes and sensors locally or in the whole heart. It also remains to be seen whether 3D-MIMs can be effectively integrated with other developing technologies, such as stem-cell bio-patches, or encapsulated viral vectors or drugs. Other challenges to the application of these membranes for chronic implantation include power supply, control, durability and encapsulation.

However, although the technology is in its infancy, the development of a 3D-MIM prototype represents a breakthrough technology with the potential to significantly expand diagnosis and treatment options for common cardiac-excitability disorders. ■

Colleen E. Clancy and Yang K. Xiang are in the Department of Pharmacology, University of California, Davis, Davis, California 95616-8636, USA. e-mail: ceclancy@ucdavis.edu

1. Xu, L. *et al.* *Nature Commun.* **5**, 3329 (2014).
2. Jia, Z. *et al.* *Circ. Arrhythm. Electrophysiol.* **4**, 753–760 (2011).
3. Zhou, X., Herbst-Robinson, K. J. & Zhang, J. *Meth. Enzymol.* **504**, 317–340 (2012).
4. Luther, S. *et al.* *Nature* **475**, 235–239 (2011).

not phosphorylated and mediates low-affinity uptake of the abundant substrate⁵ (Fig. 1). But how Thr 101 phosphorylation regulates K_m is yet to be defined. The two new studies, by Sun *et al.*¹ (page 73) and Parker and Newstead² (page 68), provide different, although not necessarily conflicting, explanations.

A transporter has three main conformational states: outward-facing (substrate site open to the external environment), occluded (closed to both sides of the membrane) and inward-facing (open to the cell cytoplasm)⁶. Using crystallography, the two groups analysed the inward-facing conformation of unphosphorylated NRT1.1, and report essentially identical structures. NRT1.1 is known to be comprised of 12 transmembrane helices (TMHs), and the researchers observed that — as with other members of the major facilitator superfamily (MFS) of membrane transporter proteins — the TMHs form a cavity open to the cytoplasm, within which the substrate

PLANT SCIENCE

How to switch affinity

The protein NRT1.1 transports nitrate ions into plants over a wide range of concentrations. Two studies provide structural insight into this unusual behaviour, but give different explanations for it. SEE ARTICLES P.68 & P.73

YI-FANG TSAY

To acquire nutrients from their surroundings, cells produce proteins that act as transporters, channels and pumps, creating passages in the plasma membrane through which molecules can pass. Unlike channels, pumps and transporters must undergo conformational changes to transfer substrates across the membrane. In plants, the transporter NRT1.1 (also known as CHL1 or NPF6.3) is essential for cellular uptake of nitrate. Unusually for a transporter, NRT1.1 regulates ion uptake by changing its affinity for nitrate ions depending on the availability of nitrate in the soil. However, the structural basis of this behaviour is not understood. Two articles^{1,2} in this issue provide possible answers, by finally determining structures of NRT1.1 two decades after its identification³.

Like enzymes, transporters exhibit an affinity for specific substrates, and their behaviour can be described by the Michaelis constant K_m , which determines the range of extracellular substrate concentrations at which a transporter can operate (a low K_m indicates that the transporter has a high affinity for a given substrate, and a high K_m indicates low affinity). The amount of nutrient available to a cell can fluctuate considerably, so, to ensure that nutrients can be acquired over a wide range of concentrations, cells have high- and low-affinity transport systems. Typically, the two are mediated by distinct transporters; however,

NRT1.1 is a dual-affinity transporter, which can switch affinity by phosphorylation and dephosphorylation at a threonine amino-acid residue (designated Thr 101)⁴.

Regulation of a transporter by phosphorylation can allow cells to rapidly respond to nutrient fluctuation. At low nitrate concentrations, NRT1.1 is phosphorylated at Thr 101, and functions as a high-affinity nitrate transporter, able to acquire scant nitrate from the soil. At high nitrate concentrations, NRT1.1 is

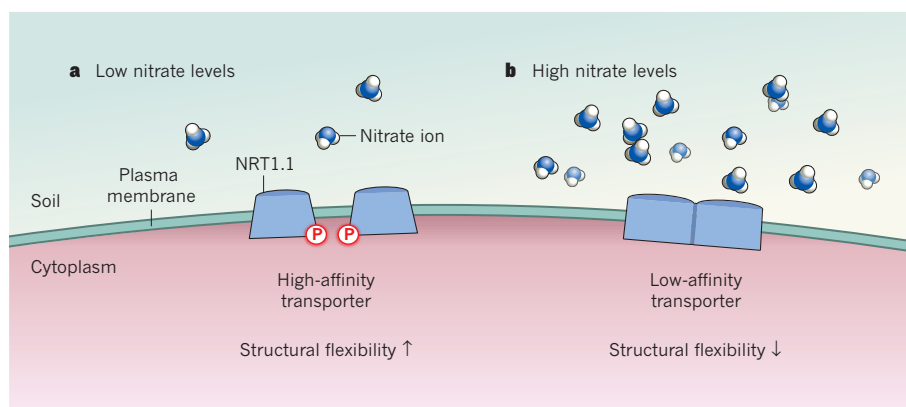


Figure 1 | Changing the affinity of the NRT1.1 protein for nitrate. The nitrate transporter NRT1.1, which spans the plasma membrane of plant cells, is responsible for the uptake of nitrate from the external environment into the cell cytoplasm. **a**, When nitrate concentration is low, NRT1.1 is phosphorylated (P) at the amino-acid residue Thr 101, giving it a high affinity for nitrate. **b**, When nitrate is abundant, Thr 101 is not phosphorylated, and NRT1.1 acts as a low-affinity transporter⁴. Sun *et al.*¹ and Parker and Newstead² have resolved the structure of unphosphorylated NRT1.1, and find that it forms a dimer. Although Sun *et al.* propose that this dimerization itself alters affinity, Parker and Newstead suggest that dephosphorylation decreases structural flexibility, modulating nitrate uptake.

can bind. Unusually for an MFS protein, NRT1.1 crystallizes as a dimer with its two monomers in the same orientation. Thr 101 is located close to the dimer interface, at the cytoplasmic end of one TMH. On the basis of data from multiple lines of analysis, Sun *et al.* propose that Thr 101 phosphorylation affects the dimerization of NRT1.1, as is seen in some transmembrane receptors. But how dimer decoupling could regulate the K_m of NRT1.1 remains undetermined.

By contrast, Parker and Newstead question the physiological significance of NRT1.1 dimerization, suggesting that it might occur only as a consequence of crystallization and not in cells. To test the function of Thr 101 phosphorylation, the authors engineered a mutant form of NRT1.1 in which Thr 101 was replaced by aspartate (forming a Thr101Asp mutant), a change that mimics permanent phosphorylation. By comparing the behaviour of Thr 101- and Thr101Asp-containing proteins, they found that the ability of nitrate to bind NRT1.1 is independent of phosphorylation.

Parker and Newstead also showed that the Thr101Asp-containing protein has a much higher transport rate (measured by hydrogen-ion uptake, which is used as a surrogate for measuring nitrate uptake) and a lower melting temperature (indicative of increased structural flexibility) than the wild-type protein. This led the authors to propose that phosphorylation increases transport rate as a result of enhanced structural flexibility, and that the change in K_m is due to this increase in the rate of nitrate transport. The seemingly disparate conclusions of the two papers can, however, be reconciled — dimer decoupling, induced by phosphorylation, might increase structural flexibility.

One caveat to Parker and Newstead's data is that it cannot be ascertained with which conformation — inward-facing or other conformational states — their biochemical measurements correlate. It is not known whether substrate affinities are identical in different conformations in MFS transporters, and so the authors cannot prove that phosphorylation does not affect the binding affinity of NRT1.1 for nitrate. To determine this, a measurement must be taken in which the protein is fixed in the outward-facing conformation, in which K_m is defined.

Both groups' structures hint that histidine residue 356 (His 356) is the nitrate-binding site. Sun and colleagues replaced His 356 with alanine, and observed that both high- and low-affinity nitrate-uptake activities were eliminated. This suggests that the His 356 nitrate-binding site is required for both high- and low-affinity transport. However, His 356 is not evolutionarily well conserved in the nitrate transporters of the NRT1 family, suggesting that other NRT1 transporters might use a different binding site for nitrate, or that there are multiple possible steps on the

nitrate-transporting path. Differences between substrate-binding sites might explain why NRT1 family members transport substrates as diverse as nitrate, peptides and hormones⁷.

The TMH regions of the independently solved structures are almost identical, but there are noticeable differences in TMH-connecting loops. These loops, especially the large central loop connecting TMH6 and TMH7 (which, unfortunately, is not well resolved in these papers) might be required for an extra function of NRT1.1: that of a nitrate receptor⁵. Sun *et al.* suggest that the TMH loops might confer receptor activity by providing docking sites at which other proteins can interact with the transporter.

The transceptor (a transporter with receptor activity) role of NRT1.1 might not be unique — the glucose transporter GLUT2 may function as a sugar transceptor to regulate insulin secretion and mediate absorption of sugar from the intestine⁸, and more transceptors are likely to be identified. In conjunction with biochemical analyses, structural analyses of NRT1.1 in other conformations, for example the less-stable phosphorylated form or in complex with other signalling components, will further our understanding of how nitrate transport and sensing are coupled and regulated.

Enhancing the efficiency with which plants take up nitrate could potentially increase crop

yields. If plants have higher nitrate uptake, this would lower the amount of nitrate that is washed off into water systems, possibly reducing the pollution of aquatic ecosystems. The reduced need for nitrogen fertilizer could also lower the world's energy burden, because production of nitrogen fertilizer accounts for 1% of global energy consumption each year. As such, development of techniques to manipulate nitrate uptake is an important agricultural issue. Resolution of the structure of NRT1.1 could now render structure-based understanding and manipulation of uptake feasible. ■

Yi-Fang Tsay is at the Institute of Molecular Biology, Academia Sinica, Taipei 11529, Taiwan.

e-mail: yftsay@gate.sinica.edu.tw

1. Sun, J. *et al.* *Nature* **507**, 73–77 (2014).
2. Parker, J. L. & Newstead, S. **507**, 68–72 (2014).
3. Tsay, Y.-F., Schroeder, J. I., Feldmann, K. A. & Crawford, N. M. *Cell* **72**, 705–713 (1993).
4. Liu, K. H. & Tsay, Y.-F. *EMBO J.* **22**, 1005–1013 (2003).
5. Ho, C. H., Lin, S. H., Hu, H. C. & Tsay, Y.-F. *Cell* **138**, 1184–1194 (2009).
6. Yan, N. *Trends Biochem. Sci.* **38**, 151–159 (2013).
7. Leran, S. *et al.* *Trends Plant Sci.* **19**, 5–9 (2014).
8. Leturque, A., Brot-Laroche, E. & Le Gall, M. *Am. J. Physiol. Endocrinol. Metab.* **296**, E985–E992 (2009).

This article was published online on 26 February 2014.

OPTOMECHANICS

Hardware for a quantum network

An optomechanical device has been designed that converts radio-frequency electrical signals into laser light. The system could allow computers to share data in a future quantum network based on optical fibres. SEE LETTER P.81

MIKA A. SILLANPÄÄ & PERTTI J. HAKONEN

Quantum computers are coming. In fact, computing technologies that are based on the peculiar laws of quantum physics are already a reality in some niche applications. These machines perform computations by manipulating quantum bits (qubits) of information, instead of classical bits. Because qubits can be 'zero' or 'one' as well as both of these values at the same time, they can encode more information than their classical analogues. Therefore, in certain applications, a quantum computer could calculate a multitude of parallel tasks much faster than supercomputers will ever be able to do. This 'qubit crunching' potential aside, quantum computers will eventually need to communicate with one another through a network. On page 81 of this issue,

Bagci *et al.*¹ describe how they have created a device that brings this goal a step closer.

The gadget devised by Bagci and colleagues uses concepts from an established field of physics called optomechanics, which studies the interaction between electromagnetic radiation and mechanical motion. When a photon strikes a mirror, it exerts a feeble kick on it, forcing the mirror to move. This effect can be boosted by placing the photon inside a cavity — an arrangement of two highly reflective mirrors that allows the photon to bounce back and forth many times between the mirrors.

The field of cavity optomechanics has made dramatic advances in the past few years². However, a challenge has been to make devices that convert radio signals into optical ones while preserving the fragile quantum information that the signals encode^{3,4}. Irrespective of how

qubit platforms in a future quantum computer will look, they will probably be operated using electrical signals of radio or microwave frequencies, because known scalable qubit designs fall in this frequency range. Optical signals can be efficiently transmitted by means of optical fibres. Therefore, such radio-to-optical converters would allow quantum processors to exchange information reliably in a network based on optical fibres. And this is where Bagci and colleagues' device comes in: the authors demonstrate a system that acts as such a converter, and that will be particularly useful for qubit platforms made from superconductors⁵.

Bagci and co-workers' system consists of a silicon nitride membrane that is coated with aluminium and so acts like a mirror. Because of its metal coating, the membrane also conducts electricity and thus can 'talk' to electrical signals. Such signals can be stored as radio-frequency photons in a cavity formed

The neat aspect of Bagci and colleagues' system is that it should be able to work in the quantum regime.

by the membrane and a planar metallic structure. By applying electrical signals to the cavity, the authors were able to set the membrane in vibrational motion. Simultaneous irradiation of the device with a laser beam

allowed the vibrations to be converted into an optical signal embedded in the laser beam as it was reflected from the mirror.

Conversion between electrical signals and laser light is a standard technique in modern communications technology; it is achieved using devices such as voltage-modulated insulating crystals, photoemitters and photodetectors. The neat aspect of Bagci and colleagues' system is that it should be able to work in the quantum regime, in which currently known techniques are likely to fail. Although the authors did not show their device operating in this domain, the system's conversion efficiency is orders of magnitude better than that of current communications technology. Once operation in the quantum regime is attained, one could imagine using the device to test features of quantum mechanics and as a component of a future quantum network.

The device also compares favourably with present-day electronic amplifiers, such as low-temperature electromechanical amplifiers⁶. In the authors' set-up, noisy electrical signals are converted into an optical beam that has a low noise component. Such low-noise amplification represents an advance over investigations of other converters^{3,4}. Hence, in addition to its potential use in quantum communication, the system may provide an alternative means of signal amplification at low-noise level in various applications that involve narrow band frequencies, such as high-resolution nuclear

magnetic resonance imaging. But there are also challenges ahead, including the need to shift from the radio to the higher, microwave frequency range, which is relevant for most potential applications, as well as for processing the output optical signal. We await further developments in this field with great interest. ■

Mika A. Sillanpää is in the Department of Applied Physics, Aalto University, FI-00076 Aalto, Finland. **Pertti J. Hakonen** is in the

O. V. Lounasmaa Laboratory, Aalto University.
e-mail: mika.sillanpaa@aalto.fi

1. Bagci, T. *et al.* *Nature* **507**, 81–85 (2014).
2. Aspelmeyer, M., Kippenberg, T. J. & Marquardt, F. Preprint at <http://arxiv.org/abs/1303.0733> (2013).
3. Andrews, R. W. *et al.* Preprint at <http://arxiv.org/abs/1310.5276> (2013).
4. Bochmann, J., Vainsencher, A., Awschalom, D. D. & Cleland, A. N. *Nature Phys.* **9**, 712–716 (2013).
5. Devoret, M. H. & Schoelkopf, R. J. *Science* **339**, 1169–1174 (2013).
6. Massel, F. *et al.* *Nature* **480**, 351–354 (2011).

INFLUENZA

Prediction is worth a shot

Predicting the long-term evolution of influenza is difficult. But a model incorporating the effects of deleterious and beneficial mutations has met the more tangible goal of predicting year-to-year frequencies of viral groups. SEE ARTICLE P.57

KATIA KOELLE & DAVID A. RASMUSSEN

Physicist Niels Bohr famously said that “prediction is difficult, especially about the future”. Perhaps even more so than in physics, this statement holds true in biology. Nevertheless, biologists have repeatedly attempted to predict how their systems will fare in the future. When it comes to systems that evolve rapidly in response to changes in their environment, prediction becomes even harder. Despite this formidable challenge, on page 57 of this issue, Łuksza and Lässig¹ tackle the problems of predicting how the influenza virus will evolve from one year to the next, and

how such prediction can improve seasonal-flu vaccines.

Certain aspects of influenza's evolutionary dynamics are, if not predictable, at least highly repetitive. Because the virus infects up to 15% of the human population each year², most individuals have some degree of immunity. However, new strains carrying mutations in epitopes (protein regions that are recognized by human antibodies) regularly arise. These strains initially have a fitness advantage over previously dominant strains because they can more effectively escape a host's immune response. As a result, they rise in frequency and, in doing so, deplete their own supply of

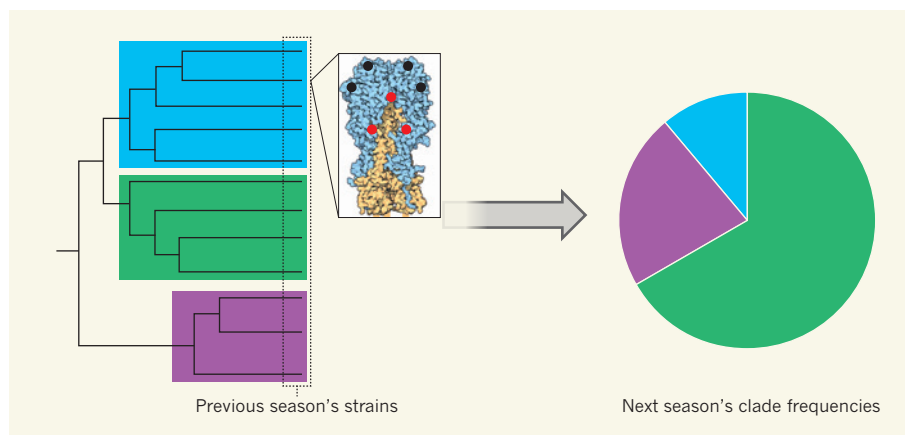


Figure 1 | Forecasting changes in influenza-clade frequencies. Łuksza and Lässig's model¹ predicts the frequencies of viral clades (blue, green and purple) in one season from the frequencies and fitness values of individual strains in the previous season. An individual strain's fitness is inferred by incorporating the beneficial effects of mutations in epitope regions of the viral haemagglutinin protein (inset, black dots) and the deleterious effects of mutations in non-epitope regions (inset, red dots). (Haemagglutinin image: David S. Goodsell/<http://doi.org/c3nqkh>.)

susceptible hosts such that even newer strains gain the advantage. This continuous process of evolution — referred to as antigenic drift — results in rapid turnover of the viral population and thus the possibility of an individual becoming reinfected with flu within a few years. Moreover, it leads to the need to regularly update the composition of seasonal-flu vaccines³.

Although flu's antigenic drift is well characterized, predicting exactly what strains carrying which antigenic mutations will circulate in the future remains problematic. This is largely because the stochastic nature of the mutational process itself leads to uncertainty over what mutations will arise. Apart from this, even predicting the fate of strains currently residing in the population is a formidable challenge, because multiple strains carrying different combinations of mutations co-circulate and to some extent compete with one another for susceptible hosts. The predictive model that Łuksza and Lässig present addresses this problem by targeting a somewhat more manageable question: can one predict changes in the frequencies of groups of viral strains (clades) from one year to the next? The answer seems to be yes, and with considerable accuracy.

At its core, Łuksza and Lässig's model predicts viral clade frequencies in a given year using strain frequencies and fitness values from the preceding year (Fig. 1). Its effectiveness therefore relies on how accurately the model assigns fitness values to strains, a difficult task to do well, given that we have little understanding of how any individual mutation affects fitness. To make this task feasible, the authors consider only the fitness effects of two classes of mutation — epitope and non-epitope mutations — in the haemagglutinin surface protein of the virus.

Although simple, this approach has a clear biological rationale. Mutations at epitopes are likely to be beneficial to the virus, because they alter the structural features targeted by host antibodies. Thus, a strain can have higher fitness than its competitors by being antigenically more distinct from previously circulating strains. By contrast, mutations outside epitope regions are often deleterious because they reduce protein stability or upset evolutionarily conserved viral functions. By training the model on the evolutionary dynamics of historical strains, the authors were able to estimate the fitness effects of these two classes of mutation, and thereby to quantify the fitness of currently circulating strains on the basis of the mutations they carried.

Using these estimates, Łuksza and Lässig projected clade frequencies a year into the future, and determined the accuracy of their model's predictions by comparing the magnitude of predicted clade-frequency changes with that of the changes observed. They found that their model correctly predicted growth in viral clades 93% of the time and correctly predicted decline 76% of the time.

The model's predictive accuracy may be very useful to the network of researchers who determine which strains to include in the seasonal-flu vaccine. Currently, vaccine strains are chosen using assays that quantify antigenic differences between circulating strains⁴. This approach is highly effective in some years, but antigenic mismatches between the vaccine strain and the strain that ends up dominating the next flu season do occur. Łuksza and Lässig's model provides insight into why such mismatches might arise: deleterious mutations in non-epitope regions might suppress the most antigenically distinct strains that are prime vaccine candidates. The model also delivers another way of choosing vaccine strains — by including strain cross-immunity estimates as well as the inferred fitness and frequency of current strains.

Although Łuksza and Lässig's model presents a new perspective on what contributes to viral fitness and what aspects of flu evolution can feasibly be predicted, it also points to necessary future work. First, the model assumes a simple relationship between the genetic distance between strains and the extent of cross-immunity that they induce in a host. However, antigenic analysis has shown that some amino-acid changes in epitope regions have only a slight antigenic effect, whereas others have a pronounced one⁵. Incorporating a more empirically informed 'map' of the associations between viral genotypes and their antigenic characteristics might increase the model's predictive power. Second, although the authors have already taken the important step of broadening predictive models to include

the effects of non-epitope mutations in the haemagglutinin protein, viral fitness also surely depends on the virus's seven other gene segments. Incorporating whole-genome analysis into the model may therefore further improve prediction, especially if mutations interact in a non-additive manner (epistatically) across gene segments.

Finally, in terms of strain selection for vaccines, it is worth bearing in mind that the ultimate goal of vaccination might not be to reduce the number of flu infections, but rather to minimize the number of flu deaths or the overall economic cost of infections⁶. Luckily, modifying the model to incorporate such aims seems relatively straightforward, provided that sufficient data are available to quantify how viral strains differ in virulence or other relevant properties. Thus, although further work is needed, it is clear that Łuksza and Lässig have significantly advanced the difficult task of flu prediction, especially about the future. ■

Katia Koelle and David A. Rasmussen are in the Department of Biology, Duke University, Durham, North Carolina 27708, USA.
e-mails: katia.koelle@duke.edu; david.rasmussen@duke.edu

1. Łuksza, M. & Lässig, M. *Nature* **507**, 57–61 (2014).
2. Stöhr, K. *Lancet Infect. Dis.* **2**, 517 (2002).
3. Carrat, F. & Flahault, A. *Vaccine* **25**, 6852–6862 (2007).
4. Russell, C. A. *et al. Vaccine* **26** (Suppl. 4), D31–D34 (2008).
5. Smith, D. J. *et al. Science* **305**, 371–376 (2004).
6. Medlock, J. & Galvani, A. P. *Science* **325**, 1705–1708 (2009).

This article was published online on 26 February 2014.

CANCER

Inflammation lights the way to metastasis

Tumour spread is the main cause of death in patients with melanoma. Exposure of melanoma to ultraviolet radiation has now been found to cause an inflammatory response that drives the formation of distant metastases. [SEE LETTER P.109](#)

SETH B. COFFELT & KARIN E. DE VISSER

Melanoma is the deadliest form of skin cancer. The main risk factor for its development is ultraviolet (UV) radiation from the sun, which directly induces alterations in the DNA of melanocytes, the skin's pigment-producing cells¹. Once a melanoma has escaped from its primary location to distant organs and formed metastases, there is only a limited chance that the cancer can be controlled. But what causes melanoma to spread? In this issue, Bald *et al.*² (page 109)

have uncovered another surprising role of UV radiation in this cancer type — it promotes metastasis through the activation of an inflammatory response.

Besides inducing mutations in melanocytes, exposure to UV radiation disrupts the epithelial cells (keratinocytes) that form the skin's outer layer, and also causes inflammation. But whether these other damaging effects of the radiation influence melanoma progression was unknown until now. Using a transgenic mouse model that closely mimics the human disease, Bald and co-workers observed

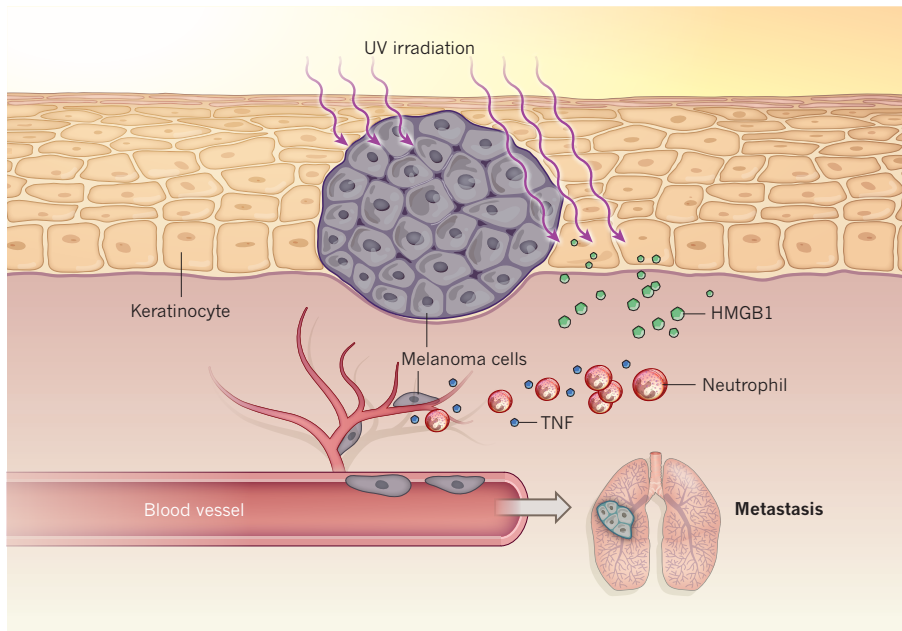


Figure 1 | Exposure to ultraviolet radiation drives metastasis. UV radiation directly induces DNA alterations that lead to melanomas. It also initiates other processes in the skin, including inflammation and damage to keratinocytes in the outer skin layer. Bald *et al.*² show that keratinocyte secretion of HMGB1 — a protein released by dying or stressed cells — initiates an influx of neutrophils. These cells produce inflammatory molecules, such as TNF, which stimulate the formation of new blood vessels and induce the migration of melanoma cells along these vessels, thereby facilitating the cancer's spread to distant organs.

that repetitive UV irradiation of developing tumours increased the formation of metastases in the lung, without affecting the growth of the primary melanoma.

Two observations led the authors to hypothesize that non-cancer cells were involved in driving metastasis. First, they noted that irradiation of melanomas caused a striking influx of neutrophils — white blood cells that are key players in immune defence and inflammatory disorders. Second, irradiation induced the migration of melanoma cells towards blood vessels and their subsequent movement along the vessels' surfaces. The authors saw that, near the tumours, the neutrophils were attracted to UV-damaged keratinocytes by a molecule called HMGB1, a protein normally found in the nucleus of healthy cells but which is released by dying or stressed cells. Depletion of neutrophils or inhibition of their HMGB1-dependent recruitment reduced the migration of melanoma cells along blood vessels and abrogated the UV-radiation-induced lung metastasis. These data indicate that the neutrophil-rich inflammatory response caused the spread of melanoma cells to distant organs (Fig. 1).

Through a series of *in vitro* experiments, the authors demonstrated that activated neutrophils secrete the pro-inflammatory protein TNF, which stimulates new blood vessel formation and the migration of melanoma cells. They also studied tumour tissues from 178 patients with melanoma and found, in line with their experimental data, that ulceration

of melanomas (breakdown of the skin layer above the cancer) and neutrophil influx were associated with the presence of melanoma cells adjacent to blood vessels and increased incidence of metastatic disease.

Thus, Bald *et al.* have uncovered a mechanism by which UV radiation triggers the activation of non-cancer cells in the vicinity of a melanoma. These cells initiate a harmful inflammatory cascade that leads to increased interaction between the cancer cells and blood vessels, culminating in the spread of the cancer cells to distant organs. The data add to the emerging realization that metastasis is not just a process that is intrinsic to cancer cells, but rather is one that depends on complex and reciprocal interactions with non-cancer cells in the tumour microenvironment³. As such, targeting UV-activated, pro-metastatic neutrophils, either by preventing their activation and accumulation or by inhibiting their downstream effects, represents a potential therapeutic opportunity to interfere with metastatic melanoma. Moreover, blocking neutrophil function may be relevant to other cancer types, because there is evidence that neutrophils can promote metastasis in other tumour models^{4,5}.

At present, it is unclear how Bald and colleagues' findings could be translated into preventive measures. Many individuals who have precancerous lesions are unaware of them, and continue to expose themselves to the sun. Furthermore, it is still uncertain how much UV-radiation exposure is required to

induce the localized cellular changes that lead to metastasis in humans. As an alternative to preventing the harmful effects of UV radiation directly (through minimizing exposure), inhibiting its downstream effector mechanisms could also be an approach to curbing metastatic melanoma. Therefore, it will be useful to determine whether TNF is the sole neutrophil-derived mediator driving UV-induced metastasis, or whether other factors are important as well. Further characterization of pro-metastatic neutrophils, for example by gene expression and functional assays, might provide greater insight into their activity. Similarly, the potential pro-metastatic effects of other cancer-cell mutations that occur through repeated exposure to UV radiation should also be determined. Such analyses will help to better define UV-triggered melanoma metastasis and may identify new targets for this deadly disease.

Previous work has shown that UV radiation not only induces melanoma-initiating mutations in melanocytes¹, but also causes a distinct mutational signature resulting in the formation of altered cellular proteins (antigens) that are recognized as foreign by the T cells of the immune system^{6,7}. This abundance of new, UV-radiation-induced antigens largely explains the recent successes seen in patients with advanced melanoma following cancer immunotherapy^{8,9}, a treatment aimed at harnessing the patient's own immune system to attack tumours. Bald and colleagues' demonstration that metastasis can be induced through the activation of keratinocytes and neutrophils is therefore the third known effect of UV radiation in melanoma. Their findings, together with those of others (reviewed in ref. 10), indicate that, in addition to activating antitumour immunity, inhibiting pro-tumour inflammatory responses is an attractive anticancer approach that might increase the number of patients with melanoma who are successfully treated. ■

Seth B. Coffelt and Karin E. de Visser are in the Division of Immunology, Netherlands Cancer Institute, 1066 CX Amsterdam, the Netherlands.
e-mail: k.d.visser@nki.nl

1. Hodis, E. *et al.* *Cell* **150**, 251–263 (2012).
2. Bald, T. *et al.* *Nature* **507**, 109–113 (2014).
3. Quail, D. F. & Joyce, J. A. *Nature Med.* **19**, 1423–1437 (2013).
4. Kowanetz, M. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 21248–21255 (2010).
5. Huh, S. J., Liang, S., Sharma, A., Dong, C. & Robertson, G. P. *Cancer Res.* **70**, 6071–6082 (2010).
6. Robbins, P. F. *et al.* *Nature Med.* **19**, 747–752 (2013).
7. van Rooij, N. *et al.* *J. Clin. Oncol.* **31**, e439–e442 (2013).
8. Wolchok, J. D. *et al.* *N. Engl. J. Med.* **369**, 122–133 (2013).
9. Couzin-Frankel, J. *Science* **342**, 1432–1433 (2013).
10. Coussens, L. M., Zitvogel, L. & Palucka, A. K. *Science* **339**, 286–291 (2013).

This article was published online on 26 February 2014.

Disclinations provide the missing mechanism for deforming olivine-rich rocks in the mantle

Patrick Cordier¹, Sylvie Demouchy², Benoît Beausir³, Vincent Taupin³, Fabrice Barou² & Claude Fressengeas³

Mantle flow involves large strains of polycrystalline aggregates. The strongly anisotropic plastic response of each individual grain in the aggregate results from the interactions between neighbouring grains and the continuity of material displacement across the grain boundaries. Orthorhombic olivine, which is the dominant mineral phase of the Earth's upper mantle, does not exhibit enough slip systems to accommodate a general deformation state by intracrystalline slip without inducing damage. Here we show that a more general description of the deformation process that includes the motion of rotational defects referred to as disclinations can solve the olivine deformation paradox. We use high-resolution electron backscattering diffraction (EBSD) maps of deformed olivine aggregates to resolve the disclinations. The disclinations are found to decorate grain boundaries in olivine samples deformed experimentally and in nature. We present a disclination-based model of a high-angle tilt boundary in olivine, which demonstrates that an applied shear induces grain-boundary migration through disclination motion. This new approach clarifies grain-boundary-mediated plasticity in polycrystalline aggregates. By providing the missing mechanism for describing plastic flow in olivine, this work will permit multiscale modelling of the rheology of the upper mantle, from the atomic scale to the scale of the flow.

There is a large body of work describing plastic deformation of olivine^{1,2}, which is by far the most abundant (~60–70%) and the weakest upper-mantle mineral under a wide range of thermo-mechanical conditions. In this literature, the prevailing deformation mechanisms include the transport of matter by diffusion and of shear by motion of dislocations. However, the dislocation-based crystal plasticity of olivine is challenged by a lack of slip systems. For plastic flow to occur homogeneously by dislocation glide alone, at least five independent slip systems must operate, according to the Von Mises criterion³. This requirement can be relaxed to four slip systems if inhomogeneous flow is allowed⁴. In orthorhombic olivine (space group *Pbnm*), plastic slip is restricted to [100] and [001] directions with no possibility of shear along [010]. Hence, only four slip systems are available: [100](010), [100](001), [001](010) and [001](100). In this case, the only non-vanishing components of the corresponding Schmid tensors (the number of overbars indicates the rank of the tensor) \bar{m} are: $m_{12}^{[100](010)} = m_{21}^{[100](010)} = 0.5$; $m_{13}^{[100](001)} = m_{31}^{[100](001)} = 0.5$; $m_{23}^{[001](010)} = m_{32}^{[001](010)} = 0.5$; and $m_{13}^{[001](100)} = m_{31}^{[001](100)} = 0.5$. Thus, the slip systems [100](001) and [001](100) are not linearly independent, and only the three-system sets ([100](010), [100](001) and [001](010)) and ([100](010), [001](010) and [001](100)) are. Therefore, olivine aggregates do not fulfil Hutchinson's relaxed condition, and arbitrary deformation can only be accommodated if additional degrees of freedom are provided, by climb (that is, dislocation motion out of the glide plane by absorption/emission of point defects) for instance. However, if climb is able to help dislocations to overcome obstacles, the climb rate is not efficient enough in olivine to provide a significant contribution to the strain (see Supplementary Information). Hirth and Kohlstedt⁵ have proposed that grain-boundary sliding accommodated by (that is, rate limited by) dislocation motion could provide an alternative mechanism and account for a creep regime

observed for olivine in which strain rate is nonlinear in stress and sensitive to grain size⁶. However, grain-boundary sliding requires displacement discontinuity⁷ and/or rotational discontinuity at grain boundaries, which is likely to induce damage. More recently, Detrez *et al.*⁸ have explored theoretically the possibility that the lack of slip systems

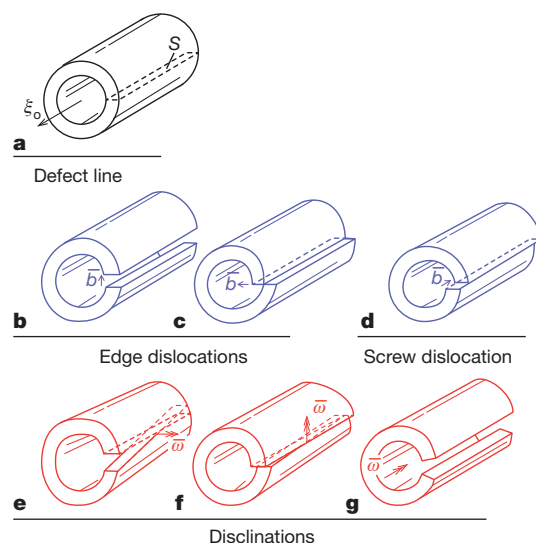


Figure 1 | Volterra's distortions. **a**, Reference cylinder with defect line ξ_0 and cut surface S . **b**, **c**, Edge dislocations with Burgers vector b . **d**, Screw dislocation with Burgers vector b . **e**, **f**, Twist disclinations with Frank vector $\bar{\omega}$. **g**, Wedge disclination with Frank vector $\bar{\omega}$.

¹Unité Matériaux et Transformations, UMR 8207 CNRS and Université Lille 1, 59650 Villeneuve d'Ascq, France. ²Geosciences Montpellier, UMR 5342 CNRS and Université de Montpellier 2, 34095 Montpellier, France. ³Laboratoire d'Etude des Microstructures et de Mécanique des Matériaux, UMR 7239 CNRS and Université de Lorraine, Ile du Saulcy, 57045 Metz Cedex, France.

Table 1 | The compatible and incompatible elasto-static defect theory

| | Translational | | Rotational |
|------------------------------------|--|---|---|
| Displacement vector | \bar{u} | Rotation vector | $\bar{\omega} = \frac{1}{2} \text{curl } \bar{u}$ |
| Strain tensor | $\bar{\varepsilon} = \frac{1}{2} (\text{grad } \bar{u} + \text{grad}^t \bar{u})$ | Curvature tensor | $\bar{\kappa} = \text{grad } \bar{\omega}$ |
| Cauchy stress tensor | $\bar{\sigma} = \bar{\mathbb{C}} : \bar{\varepsilon}^e$ | Couple-stress tensor | $\bar{\mathbb{M}} = \bar{\mathbb{A}} : \bar{\kappa}^e$ |
| Equilibrium condition for stresses | $\text{div } \bar{\sigma} = 0$ | Equilibrium condition for couple stresses | $\text{div } \bar{\mathbb{M}} = 0$ |
| Nye's dislocation density tensor | $\bar{\alpha} = \text{curl } \bar{U}_e = -\text{curl } \bar{U}_p$ | Disclination density tensor | $\bar{\theta} = \text{curl } \bar{\kappa}_e = -\text{curl } \bar{\kappa}_p$ |

The left column displays the features of translational elasticity and plasticity; the right column shows the rotational counterparts. The derivation of these elements is provided in the Supplementary Information.

in olivine could be overcome by purely diffusive (that is, linear) mechanisms operating at grain boundaries. They showed, however, that such a mechanism was unable to sustain a nonlinear rheology.

Disclinations in solids

Here we show that the lack of slip systems in olivine can be offset by a more general analysis of plastic deformation in solids. Indeed, plastic deformation does not result only from the motion of dislocations. Dislocations are the crystal defects arising from translational lattice incompatibility, as measured by the Burgers vector or Nye's dislocation density

tensor $\bar{\alpha}$ (see Supplementary Information for a summary of the elasto-plastic field theory of crystal defects). Similarly, the rotational incompatibility of the crystal lattice can be related to defects called disclinations. Dislocations and disclinations were both proposed by Volterra⁹ to account for the discontinuity of elastic displacements and rotations along surfaces in a solid containing defects (Fig. 1). In the thought experiment proposed by Volterra, a cut is made in a defect-containing elastic cylinder, from which the core of the line defect, lying along the cylinder axis, has been removed. In the presence of lattice incompatibility, the cut induces rigid-body motion of one edge of the cut with

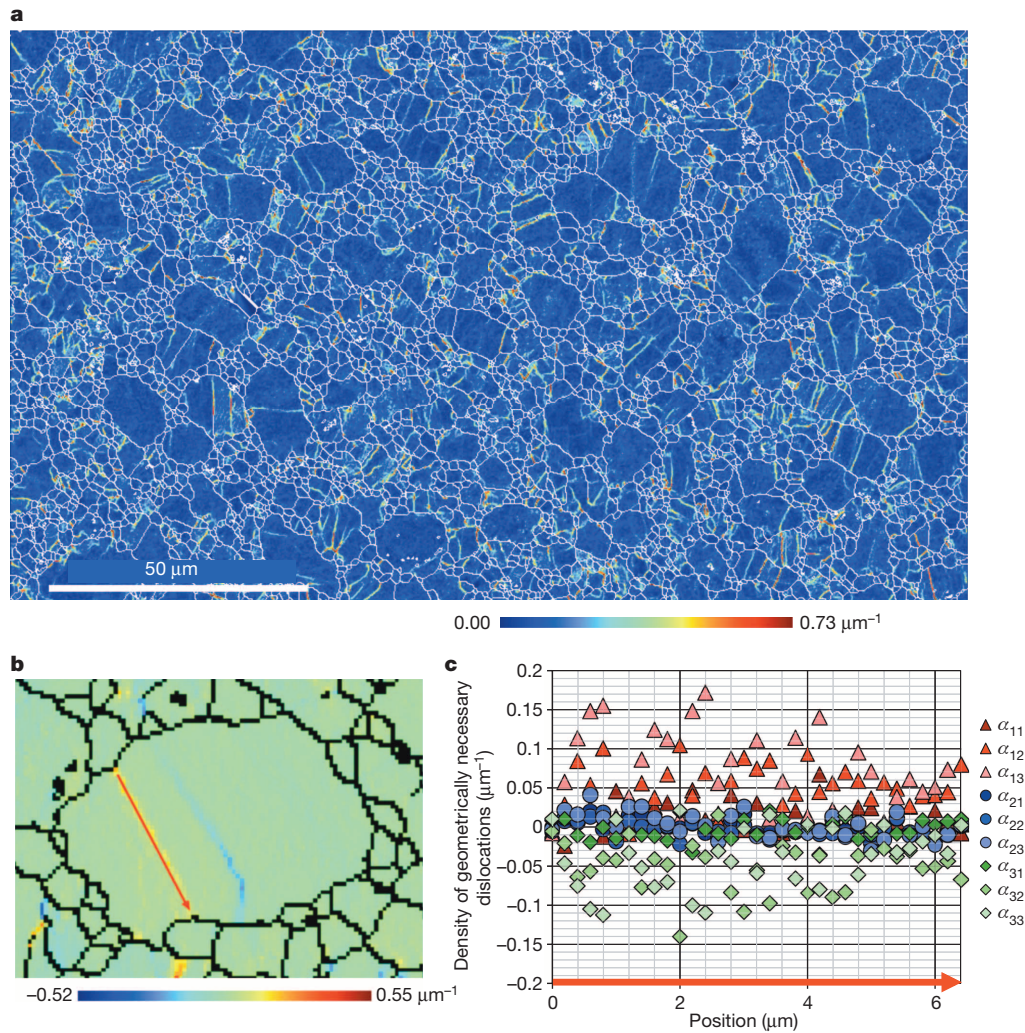


Figure 2 | Geometrically necessary dislocation densities in olivine. Densities are given per micrometre. **a**, Entrywise norm of Nye tensor $\bar{\alpha}$ in sample PoEM22 (experimentally deformed at 8% in compression at 900 °C). **b**, Close-up, showing the α_{13} component with an opposite curvature on the subgrain boundary. **c**, Nye tensor components projected in the crystal reference system

along a subgrain boundary (red arrow); the [100]-glide, [010]-glide and [001]-glide are plotted with red, blue and green marks, respectively. With [100] and [001] edge dislocations prevailing, the subgrain boundary is essentially a tilt boundary. We note the occasional presence of a twist component through sets of screw dislocations at a right angle (α_{11}, α_{33}).

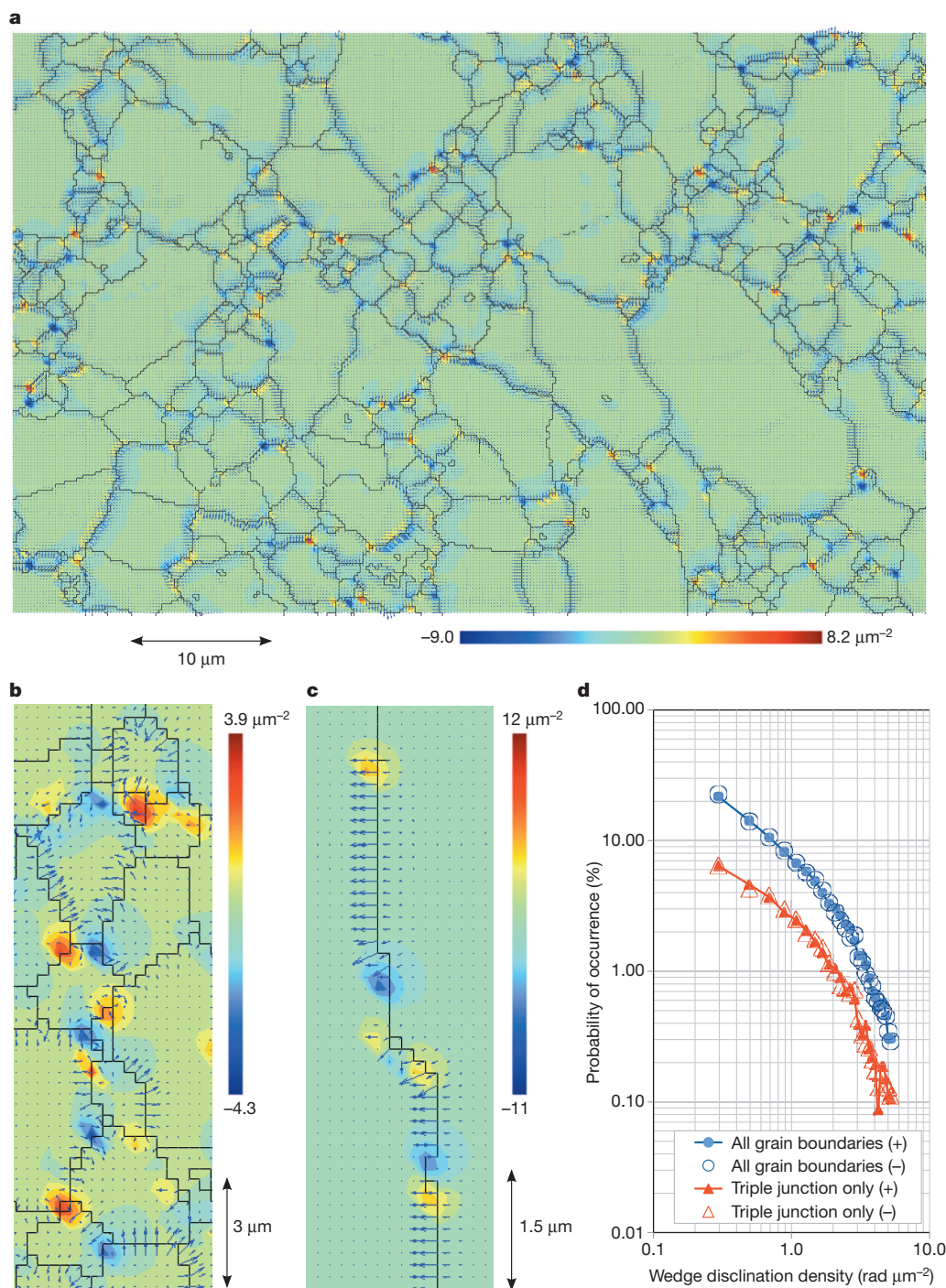


Figure 3 | Three maps representing the density of wedge disclinations θ_{33} in three deformed olivine aggregates, and the probability of occurrence for sample T0548. The density is given in radians per square micrometre. (See below and Supplementary Information for a detailed description of the samples.) The local Burgers vectors arising from edge dislocations are represented by the blue arrows: their horizontal and vertical components are respectively α_{13} and α_{23} (given per micrometre). **a**, Sample PoEM22 (experimentally deformed 8% compression at 900 °C). This low-magnification

map shows the pervasive occurrence of disclinations in grain boundaries.

b, Sample T0548 (experimentally deformed 250% torsion at 1,200 °C).

c, Sample OOM, a naturally deformed mylonitic harzburgite from the Oman

ophiolite (Sumail massif). **d**, Probability of occurrence of positive (+) versus negative (−) wedge disclination density θ_{33} in sample T0548. Filled circles indicate all positive wedges, open circles indicate all negative wedges, filled triangles indicate positive triple junction wedges and open triangles indicate negative triple junction wedges.

respect to the other. When this motion is a translation, the defect is referred to as a dislocation, whose strength is the so-called Burgers vector, that is, the (space-independent) translation vector. A disclination is obtained when the motion is a pure rotation. The strength of the disclination is the relative rotation vector of the undeformed edges of the cut, referred to as the Frank vector. The relative displacement of the

edges in this rotation also gives rise to a space-dependent Burgers vector associated with the disclination. As rotational defects, disclinations are set into motion by moments of stresses. This motion results in disclination-mediated plasticity not accounted for by the dislocation theory. Disclinations have long been neglected in the field theory of crystal defects, including for deep Earth minerals, owing to the very large level of elastic

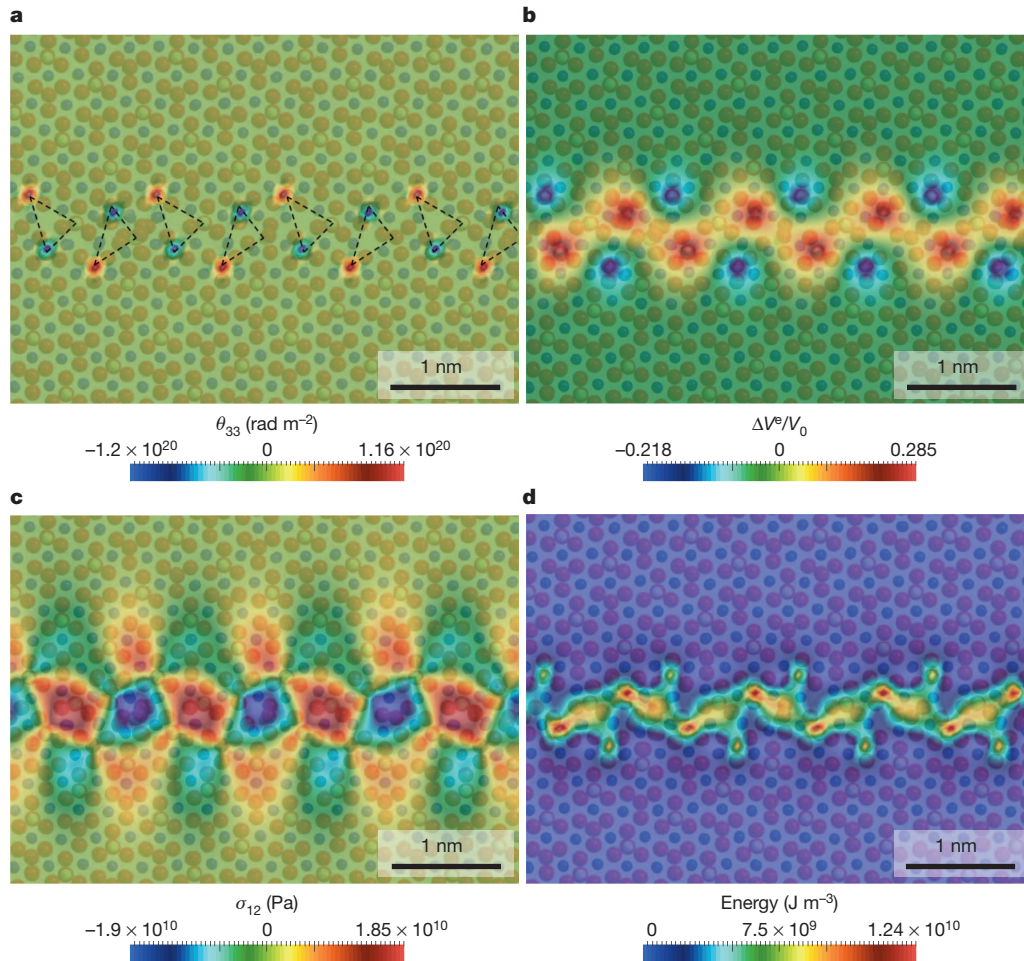


Figure 4 | Disclination-based modelling of the (011)/[100] tilt grain boundary with misorientation 60° modelled at the atomic scale in olivine²¹. **a**, Initial wedge disclination density θ_{33} (given in radians per square metre) used in the simulations. The triangles represent the structural units

energy they involve, compared with dislocations, which precludes their occurrence as isolated crystalline objects¹⁰. However, self-screened configurations, such as disclination dipoles, involve relatively small elastic energy levels^{11,12}. It was recently shown that disclination dipoles are pervasive along boundaries and subgrain boundaries in severely deformed copper, inclusion-free steel, electrodeposited aluminium thin film and in recrystallized titanium¹³, highlighting the importance of rotational defects in metallic polycrystals.

Disclinations evidenced by EBSD

To check whether, like metals, deformed olivines contain disclinations, we used EBSD, which is a technique well adapted to local lattice orientation measurements. From the orientation differences $\Delta\theta_i$ between neighbouring points separated by Δx_j the elastic curvature tensor (see Table 1) can be captured:

$$\kappa_{ij}^e = \frac{\partial\theta_i}{\partial x_j} \approx \frac{\Delta\theta_i}{\Delta x_j} \quad (1)$$

In fact, only six components of the elastic curvature tensor can be determined, because differences along the direction perpendicular to the surface are not available¹⁴ (see Methods). The Nye's dislocation density tensor can then be approximately derived as

$$\alpha_{ik} = \delta_{ik} \kappa_{mm}^e - \kappa_{ki}^e \quad (2)$$

Using this equation and considering the constraints on the measurements, five dislocation densities can be recovered in the reference frame

proposed by ref. 21 (see Methods for the construction of this model). **b**, Elastic dilations and contractions. **c**, Shear stress σ_{12} (given in pascals). **d**, Elastic energy density (given in joules per cubic metre).

of the laboratory, namely α_{12} , α_{13} , α_{21} , α_{23} and α_{33} , if the surface lies along the (1, 2) directions. Knowing the crystal orientation at each measurement point, these dislocation densities can be rotated in the crystallographic reference frame (they are then referred to as α_{ij}^c , where 'c' indicates 'crystallographic'). Figure 2a, b shows a scalar dislocation measure (that is, the length of the local Burgers vector per unit surface resulting from the edge dislocation densities α_{13} , α_{23}) in olivine polycrystals deformed experimentally^{15,16}. One of the most remarkable features of these maps is the occurrence of numerous straight subgrain boundaries, especially in the larger grains. These subgrain boundaries are commonly observed in olivine grains^{17–19} in naturally deformed mantle rocks. Made of geometrically necessary dislocations²⁰, subgrain boundaries accommodate intracrystalline misorientations and represent a first indication (largely overlooked) of the importance of these components in the plastic deformation of olivine. The analysis of the α_{ij}^c components allows us to characterize the dislocation types present in the sample. Figure 2c shows such an analysis performed along one of the subgrain boundaries (Fig. 2b). The boundary is made of [100] dislocations ($\alpha_{11}^c, \alpha_{12}^c, \alpha_{13}^c$) described in the crystal reference frame and of [001] dislocations ($\alpha_{31}^c, \alpha_{32}^c, \alpha_{33}^c$). The components corresponding to [010]-glide ($\alpha_{21}^c, \alpha_{22}^c, \alpha_{23}^c$) are negligible, in agreement with the known characteristics of olivine. Further characterization within the samples considered in this study leads to the same conclusion and demonstrates that our analysis allows the dislocation content in deformed olivine to be resolved. Previously, only decoration (annealing followed by optical microscopy or scanning electron microscopy) or transmission electron microscopy

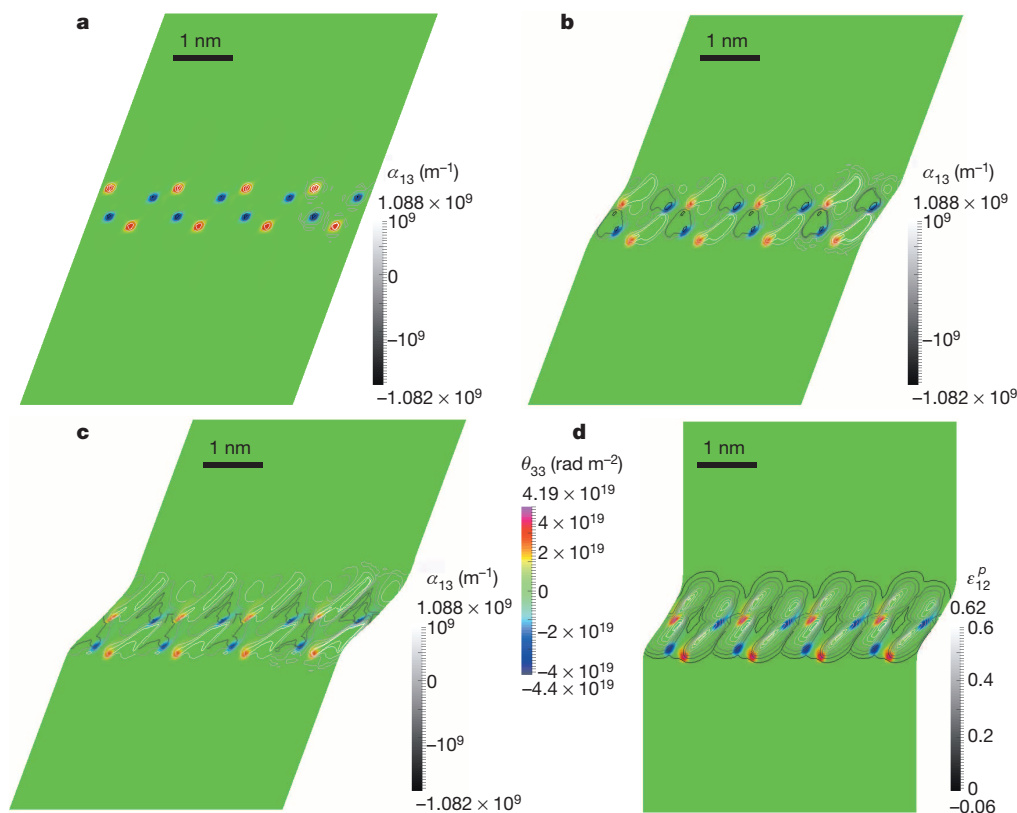


Figure 5 | Shear-coupled boundary migration of the (011)/[100] tilt grain boundary of Fig. 4. The wedge disclination density θ_{33} (in radians per square metre) is colour-coded. A to C shows the downward migration of the disclinations from 0 ms to 170 ms, when a positive shear strain of 0.2 is applied. Black and white contours show the edge dislocation density α_{13} (per metre). D shows the unloaded bicrystal after 170 ms, revealing shear-coupled migration. Black and white contours show the positive plastic shear strain ε_{12}^p produced by the downward migration of the grain boundary. For this particular misorientation, the shear coupling factor $\beta = -0.8$ resulting from a downward migration coupled with positive shear is negative, but it may be positive for different misorientations.

were available to characterize dislocation microstructures in olivine. This new technique also has another use, in that the variations in space of the elastic curvatures tensor yield the disclination densities:

$$\theta_{ij} = e_{jkl} \kappa_{il,k}^e \quad (3)$$

Figure 3 presents the results of this analysis performed on the experimentally and naturally deformed samples. It shows that the grain boundaries are associated with dislocation densities but also that they contain numerous disclinations correlating with the variations in misorientation along the grain boundary. Disclinations are also frequently associated with triple junctions. A visual inspection suggests that positive and negative disclinations are spatially coupled in the form of dipoles. The statistical analysis of Fig. 3d demonstrates that positive and negative wedge disclination densities balance perfectly, not only for the set of all grain boundaries but also for the subset composed of triple junctions. The present study shows that disclination dipoles and the rotational incompatibility they reveal represent a significant component of the deformation field in plastically deformed olivines, which is localized in grain boundaries and complements the known contribution of dislocations.

Disclinations-based grain-boundary migration

To illustrate the role of disclinations on plasticity, we built a model of a tilt boundary in olivine with a periodic wedge disclination array. We started from the (011)/[100] tilt grain boundary with misorientation 60° already modelled at the atomic scale in olivine²¹. Figure 4 shows the disclination model built on this boundary. The wedge disclination density spots are located on the vertices of the structural units²¹. The zigzag arrangements of the wedge disclinations result in a very efficient self-screening quasi-quadrupole configuration¹¹. Indeed, most of the energy is located within the structural units identified on the atomistic model. Our disclination-based model leads to a grain-boundary energy of 1.3 J m^{-2} , in excellent agreement with the value proposed by Adjaoud *et al.*²¹ based on atomistic calculations of 1.28 J m^{-2} . When a shear stress is applied to this disclination distribution, the strong variations of the local shear strain within the defect-containing areas generate couple stresses. In response to this couple-stress field, the disclination

dipoles are set into motion normal to the boundary (Fig. 5), producing plastic shear parallel to the boundary. The disclination dipole structure is maintained during this motion, which induces both migration of the boundary and shear of the crystal. Like other grain-boundary deformation mechanisms, grain-boundary migration/disclination motion is dependent on grain size. It is more apparent when the proportion of matter involved in grain boundaries is higher, that is, at small grain sizes.

We note that this disclination-induced grain-boundary migration mechanism can exhibit linearity, or nonlinearity in stress, depending on the assumed relationship between disclination velocity and the associated driving force. A more comprehensive study of the disclination structure, energy and mobility of grain boundaries as functions of misorientations is clearly needed to reach an eventual description of the rheology of an olivine aggregate. EBSD-based disclination imaging in naturally deformed olivine aggregates will allow these models to be established. Our model might also explain the formation of crystal-preferred orientations in materials deformed in the Newtonian creep regime²², as recently documented for olivine²³.

METHODS SUMMARY

We used high-resolution EBSD to analyse the microstructure of four aggregates. These were a mylonitic harzburgite (OOM) from the Oman ophiolite (Sumail massif) and three olivine aggregates: a hot-pressed undeformed sample (PI-1619; ref. 15), a deformed sample in compression at 300 MPa and 900°C (PoEM22) and a deformed sample under torsion at 300 MPa and $1,200^\circ \text{C}$ (T0548). EBSD analyses were conducted with a CamScan X500FE CrystalProbe. The operating conditions were 15 kV and 2.5-nA current under low-vacuum conditions (4–5 Pa of gaseous nitrogen) with a 20-mm working distance. The step size was $0.15 \mu\text{m}$ or $0.2 \mu\text{m}$. The data were acquired and treated with CHANNEL5 software (http://caf.ua.edu/wp-content/uploads/docs/JEOL-7000F-Oxford_Channel_5_User_Manual.pdf).

The elastic curvature tensor $\bar{\kappa}_e^i$ (where 'e' indicates 'elastic') was recovered from the EBSD orientation maps. By additionally recovering the elastic strain field $\bar{\varepsilon}_e$ and building the curl of this field $\text{curl } \bar{\varepsilon}_e$, the dislocation density tensor $\bar{\alpha}$ can be inferred exactly. In this work, as in most studies, the $\text{curl } \bar{\varepsilon}_e$ term is overlooked, and $\bar{\alpha}$ is approximated by

$$\bar{\alpha} = \text{Tr}(\bar{\kappa}_e^i) \bar{I} - \bar{\kappa}_e^i \quad (4)$$

In addition, the disclination density tensor $\bar{\theta}$ may be recovered by building the curl of $\bar{\kappa}_c$. When a single planar orientation map is known, only five components of $\bar{\kappa}$ and three components of $\bar{\theta}$ can be determined. If the planar directions are labelled (1, 2) in the sample frame, the available components are $(\alpha_{12}, \alpha_{13}, \alpha_{21}, \alpha_{23}, \alpha_{33})$ and $(\theta_{13}, \theta_{23}, \theta_{33})$. To predict the grain-boundary structure and mobility, we used a theory of crystal defect fields (disclinations and dislocations) defined at the interatomic scale^{24,25}. Our modelling paradigm is to account for the lattice incompatibility arising from crystal defects by focusing on the defect densities, rather than on the atoms themselves. The theory has the standard mathematical structure of a set of partial differential equations with boundary conditions. The unknown fields are the tensorial defect densities (dislocations and disclinations) and displacement vector fields, with standard boundary conditions on displacements and traction/moment vectors. As a consequence of this boundary value structure, approximate solutions can be generated by using finite-element methods²⁶.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 July 2013; accepted 14 January 2014.

Published online 26 February 2014.

- Kohlstedt, D. L., Evans, B. & Mackwell, S. J. Strength of the lithosphere—constraints imposed by laboratory experiments. *J. Geophys. Res.* **100** (B9), 17587–17602 (1995).
- Hirth, G. & Kohlstedt, D. L. in *Inside the Subduction Factory* (ed. Eiler, J.) 83–105 (American Geophysical Union, 2003).
- Von Mises, R. Mechanik der plastischen Formänderung von Kristallen. *Z. Angew. Math. Mech.* **8**, 161–185 (1928).
- Hutchinson, J. W. Creep and plasticity of hexagonal polycrystals as related to single crystal slip. *Metall. Trans. A* **8**, 1465–1469 (1977).
- Hirth, G. & Kohlstedt, D. L. Experimental constraints on the dynamics of the partially molten upper-mantle. 2. Deformation in the dislocation creep regime. *J. Geophys. Res.* **100** (B8), 15441–15449 (1995).
- Hansen, L. N., Zimmerman, M. E. & Kohlstedt, D. L. Grain boundary sliding in San Carlos olivine: flow law parameters and crystallographic-preferred orientation. *J. Geophys. Res.* **116**, B08201 (2011).
- Langdon, T. G. Grain boundary sliding revisited: developments in sliding over four decades. *J. Mater. Sci.* **41**, 597–609 (2006).
- Detrez, F., Castelnau, O., Cordier, P., Merkel, S. & Raterron, P. Second-order theory for the effective behavior in viscoplastic polycrystals without sufficient slip system families: application to olivine. *J. Mech. Phys. Solids* (submitted).
- Volterra, V. Sur l'équilibre des corps élastiques multiplement connexes. *Ann. Sci. Ecole Norm. Supér.* **24**, 401–517 (1907).
- Friedel, J. *Dislocations* (Pergamon, 1967).
- Romanov, A. E. & Vladimirov, V. I. in *Dislocations in Solids* Vol. 9 (ed. Nabarro, F. R. N.) 191–402 (North-Holland, 1992).
- Romanov, A. E. & Kolesnikova, A. L. Application of disclination concept to solid structures. *Prog. Mater. Sci.* **54**, 740–769 (2009).
- Beausir, B. & Fressengeas, C. Disclination densities from EBSD orientation mapping. *Int. J. Solids Struct.* **50**, 137–146 (2013).
- Pantleon, W. Resolving the geometrically necessary dislocation content by conventional electron backscattering diffraction. *Scr. Mater.* **58**, 994–997 (2008).
- Demouchy, S., Tommasi, A., Barou, F., Mainprice, D. & Cordier, P. Deformation of olivine in torsion under hydrous conditions. *Phys. Earth Planet. Inter.* **202**, 56–70 (2012).
- Seront, B. *Déformation Expérimentale à Haute Pression et Haute Température d'Agrégats Polycristallins de Plagioclase et d'Olivine*. PhD thesis, Montpellier Univ. (1993).
- Arndt, N. T. et al. Olivine, and the origin of kimberlite. *J. Petrol.* **51**, 573–602 (2010).
- Soustelle, V., Tommasi, A., Demouchy, S. & Ionov, D. A. Deformation and fluid-rock interaction in the supra-subduction mantle: microstructures and water contents in peridotite xenoliths from the Avacha Volcano, Kamchatka. *J. Petrol.* **51**, 363–394 (2010).
- Falus, G., Tommasi, A. & Soustelle, V. The effect of dynamic recrystallization on olivine crystal preferred orientations in mantle xenoliths deformed under varied stress conditions. *J. Struct. Geol.* **33**, 1528–1540 (2011).
- Ashby, M. F. Deformation of plastically non-homogeneous materials. *Phil. Mag.* **21**, 399–424 (1970).
- Adjaoud, O., Marquardt, K. & Jahn, S. Atomic structures and energies of grain boundaries in Mg_2SiO_4 forsterite from atomistic modeling. *Phys. Chem. Miner.* **39**, 749–760 (2012).
- Gomez Barreiro, J. et al. Preferred orientation of anorthite deformed experimentally in Newtonian creep. *Earth Planet. Sci. Lett.* **264**, 188–207 (2007).
- Miyazaki, T., Sueyoshi, K. & Hiraga, T. Olivine crystals align during diffusion creep of Earth's upper mantle. *Nature* **502**, 321–326 (2013).
- Fressengeas, C., Taupin, V. & Capolungo, L. An elasto-plastic theory of dislocation and disclination fields. *Int. J. Solids Struct.* **48**, 3499–3509 (2011).
- Taupin, V., Capolungo, L., Fressengeas, C., Das, A. & Upadhyay, M. Grain boundary modeling using an elasto-plastic theory of dislocation and disclination fields. *J. Mech. Phys. Solids* **61**, 370–384 (2013).
- Varadhan, S. N., Beaudoin, A. J., Acharya, A. & Fressengeas, C. Dislocation transport using an explicit Galerkin/least-squares formulation. *Model. Simul. Mater. Sci. Eng.* **14**, 1245–1270 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge financial support from the European Research Council under the Seventh Framework Programme (FP7—ERC grant number 290424—RheoMan), from a Marie Curie fellowship (FP7-PEOPLE-20074-3-IRG, grant number 230748-PoEM) and from the Agence Nationale de la Recherche (grant number ANR-11-JS09-007-01, NanoMec).

Author Contributions S.D. deformed the olivine samples and performed the EBSD measurements (with the help of F.B.). B.B., V.T. and C.F. performed the data analysis and disclination modelling. P.C. wrote the paper with feedback and contributions from all co-authors. All authors discussed and interpreted the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (patrick.cordier@univ-lille1.fr).

A predictive fitness model for influenza

Marta Luksza^{1,2} & Michael Lässig¹

The seasonal human influenza A/H3N2 virus undergoes rapid evolution, which produces significant year-to-year sequence turnover in the population of circulating strains. Adaptive mutations respond to human immune challenge and occur primarily in antigenic epitopes, the antibody-binding domains of the viral surface protein haemagglutinin. Here we develop a fitness model for haemagglutinin that predicts the evolution of the viral population from one year to the next. Two factors are shown to determine the fitness of a strain: adaptive epitope changes and deleterious mutations outside the epitopes. We infer both fitness components for the strains circulating in a given year, using population-genetic data of all previous strains. From fitness and frequency of each strain, we predict the frequency of its descendent strains in the following year. This fitness model maps the adaptive history of influenza A and suggests a principled method for vaccine selection. Our results call for a more comprehensive epidemiology of influenza and other fast-evolving pathogens that integrates antigenic phenotypes with other viral functions coupled by genetic linkage.

The evolution of influenza A/H3N2 is well documented by sequence data of several thousand strains since 1968¹. Most of these data contain the gene sequence of haemagglutinin (HA), which covers one of eight segments of the influenza genome and is the primary locus of interaction with the human immune system². Consistent with this functional role, antigenic changes in the HA epitopes carry the adaptive evolution of the pathogen^{3–11}.

Evolutionary analysis has a particular role for influenza: it serves not only to reconstruct the dynamical process and its causes, but to predict future changes^{3,4}. Any prediction of evolution is essentially an estimate of fitness differences between strains. It is these differences that lead to deterministic changes in population frequency, which are predictable if we know how fitness depends on genotype and host environment. Predictability is limited by stochastic events, which range from mutations in individual viral sequences to sampling in host-to-host transmission. Predictions of influenza HA evolution can inform vaccine selection if, despite this limitation, they are sufficiently accurate from one year to the next. Currently, the selection of vaccine strains is based primarily on haemagglutination inhibition assays, which are used to map antigenic changes between viral strains¹². But the fitness of a strain is a complex phenotype, which integrates antigenic properties with multiple other molecular functions, one of which is simply the thermodynamic stability of proteins^{13,14}. Because there is no recombination, the evolution of these functions is strongly coupled, at least within each genomic segment^{9,10} (whereas genetic linkage between segments is reduced by reassortment¹⁵). Here we show that this coupled dynamics can be captured by a fitness model that predicts the evolution of influenza from genomic data.

Clades as units of prediction

Our analysis is based on a sample of 3,944 unique HA coding sequences obtained from influenza A/H3N2 isolates between 1968 and 2012 (ref. 1), partitioned into half-year seasons (Methods). The HA sequences of a given season differ from each other by several epitope and non-epitope nonsynonymous point mutations. To quantify this diversity, we can estimate the population frequencies of mutant alleles at individual RNA sites, of combinations of mutant alleles at two or more sites, and of individual strains. From an epidemiological point of view, the frequency of a strain is simply the fraction of the infected host individuals corresponding to that strain¹⁶. We infer the genealogy of these strains by an

ensemble of trees; see Methods for details of frequency estimation and tree reconstruction. We can then trace the evolution of strain lineages or clades, which are defined as sets of strains descending from a common ancestor (Fig. 1). Whereas strains are typically observed only in a single season, clades have an evolutionary history that extends up to about 5 years and ends with fixation or loss⁷. Clades destined for fixation originate on the so-called trunk of the tree; all other clades are destined

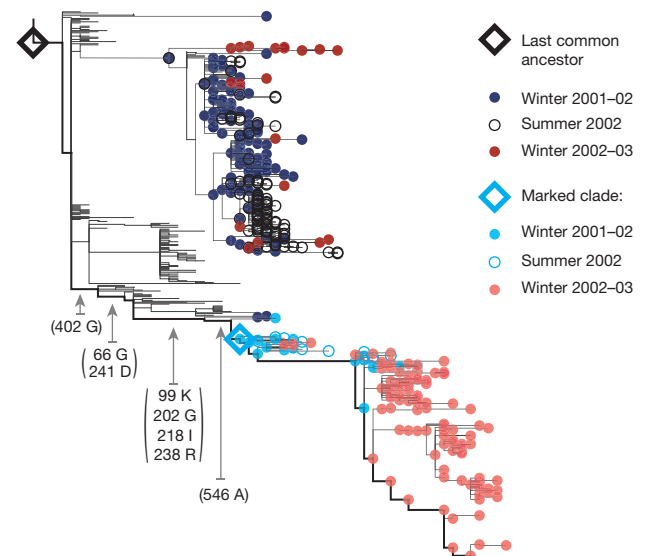


Figure 1 | Evolution of influenza clades. The figure shows a partial influenza strain tree, which is based on strains observed in years 2002 and 2003 (bullets and circles). Each strain i has a frequency x_i in its season's strain population. Our units of prediction are clades, which are defined as sets of strains descending from recent last common ancestors. For one of these clades, we mark its strain content in winter seasons $t = 2002$ and $t + 1$ (light-colour bullets) and its last common ancestor (blue diamond). Each clade is linked by a set of mutations to the last common ancestor of all strains in year t (black diamond); codon position and target amino acid of these mutations are indicated for the marked clade. A clade v observed in season t has a frequency $X_v(t)$, which is the sum of the frequencies of its strains in season t . The marked clade grows substantially from $X_v(t) = 0.08$ to $X_v(t + 1) = 0.86$.

¹Institute for Theoretical Physics, University of Cologne, Zùlpicher Strasse 77, 50937 Köln, Germany. ²Biological Sciences, Columbia University, 607D Fairchild Center, New York, New York 10027, USA.

for loss (Fig. 1). The evolution of these clades is what we want to predict from one year to the next. A successful clade diversifies from its ancestor strain through subsequent mutations during its expansion in the population. At the same time, the same mutation often originates independently in different clades. It is specific combinations of mutations that distinguish each clade from the other coexisting clades. We make predictions for these clades by averaging over the ensemble of equiprobable trees, which minimizes the effects of tree reconstruction ambiguities³ (Methods).

Our prediction is based on frequency and fitness data that depend only on information actually available at a given point in time. Consider a clade v containing a set of strains i with frequencies x_i in a given season t . The observed frequency of that clade in season t , which is denoted as $X_v(t)$, is simply the sum of these strain frequencies, $X_v(t) = \sum_{i \in v, t} x_i$. This sum is defined as an average over strain trees, as detailed in Methods. Each strain has a Malthusian fitness or growth rate f_i (measured in units of 1/year), which is to be specified by our model. Given these initial data, we predict the frequency of that clade in the season 1 year later,

$$\hat{X}_v(t+1) = \sum_{i \in v, t} x_i \exp(f_i) \quad (1)$$

as illustrated in Fig. 1 (for details, see Methods). Equation (1) describes the large-scale population dynamics averaged over many transmission cycles and over the yearly epidemic cycle. We restrict predictions to clades with frequencies $X_v(t) > 0.15$, which are large enough for reliable estimation. These clades are geographically well-mixed^{17,18} (93% of them cover two or more continents), whereas smaller clades are dominated by sampling noise and geographical bias (94% are observed on a single continent only). We can check the quality of our method a posteriori by comparing predicted and actual clade evolution, using the observed frequencies $X_v(t+1)$.

Fitness model

Our fitness model has two components, which describe the selection on epitope and non-epitope HA genotypes, respectively. Epitope changes are predominantly under positive selection^{7–11}, because they affect the antigenic characteristics of a strain. Antigenic selection is contained in multi-strain epidemiological models, which describe a susceptible–infected–recovered (SIR) dynamics^{19–23}. In this type of model, host individuals acquire partial immunity against infections with all strains of similar antigenic characteristics. Therefore, the strain growth rates f_i depend on the population history of previously circulating strains. We use an SIR model to derive our minimal epitope fitness model (Methods): a given strain i incurs a cross-immunity load generated by all previous strains j , each of which generates a fitness cost proportional to its frequency x_j and to the cross-immunity amplitude $\mathcal{C}(\mathbf{a}_i, \mathbf{a}_j)$. These amplitudes depend on the antigenic similarity of the strains i and j , which is encoded in the epitope segments of their HA sequences \mathbf{a}_i and \mathbf{a}_j . We neglect higher-order antigenic interactions involving more than two strains¹⁹ and the birth–death turnover of the host population, which can be argued to produce only subleading effects in the epidemiology of influenza A/H3N2.

Non-epitope mutations are predominantly under negative selection⁹, because they affect protein stability and other conserved molecular functions^{13,14}. Here we describe these effects by a simple mutational-load model: each strain incurs a fitness cost $\mathcal{L}(\mathbf{a}_i)$ that is the cumulative effect of recent non-epitope amino acid changes, which occur in its ancestral lineage in the current season (Methods).

Together we obtain a strain fitness of the form

$$f_i = f_0 - \mathcal{L}(\mathbf{a}_i) - \sum_{j: t_j < t_i} x_j \mathcal{C}(\mathbf{a}_i, \mathbf{a}_j) \quad (2)$$

with a constant f_0 ensuring the correct normalization of strain frequencies (Methods). Importantly, this strain-based model goes beyond a fitness model for individual mutations: it counts each new beneficial or

deleterious mutation together with the previous changes in its ancestral lineage.

The simplest fitness model of this form has uniform selective effects: each non-epitope mutation generates a fitness cost σ_{ne} , and each epitope mutation reduces the cross-immunity amplitude by an amount σ_{ep} . However, the biology of cross-immunity and protein stability deviates from this model. Both phenotypes are non-uniform and nonlinear functions of genetic distance^{4,12–14,24}; that is, the effect of a mutation depends on its sequence position, on the amino acids involved, and on its background of previous mutations. Our full fitness model uses nonlinear cross-immunity amplitudes $\mathcal{C}(\mathbf{a}_i, \mathbf{a}_j)$, and it includes position-specific effects and nonlinear fitness terms that are inferred from observed clade histories (Methods; see also a related allele-based inference scheme¹⁰). Importantly, this model has only four fit parameters, which can be inferred from our data set without compromising predictive power.

Frequency predictions for clades

The winter-to-winter prediction for the Northern Hemisphere obtained from our full, clade-based fitness model is shown in Fig. 2. To determine the accuracy of this model, we compare the predicted frequency ratio, or Wrightian fitness, $\hat{W}_v = \hat{X}_v(t+1)/X_v(t)$ with the posterior observed ratio $W_v = X_v(t+1)/X_v(t)$ for all clades with frequencies $X_v(t) > 0.15$ in a given season. The data points (W_v , \hat{W}_v) in Fig. 2a are distributed around the diagonal of correct prediction; some scatter can be explained by statistical errors in frequency estimation due to tree reconstruction and sampling (Extended Data Fig. 1). As discussed below, these predictions can be improved further by broadening the data basis of our model. The direction of frequency evolution is predicted with remarkable accuracy. There are 121 clades with observed growth ($W_v > 1$), which we predict correctly in 93% of the cases ($\hat{W}_v > 1$). For the 67 clades with observed decline ($W_v < 1$), we correctly predict decline in 76% of the cases. Importantly, the fitness amplitude W_v is predicted accurately for the clades destined for fixation, which have $\hat{X}_v(t+1) \approx 1$ and appear close to the diagonal in Fig. 2a.

The fixation of a clade implies the fixation of all mutations that appear in its ancestor strain. As shown in Fig. 2b, the yearly numbers of nucleotide fixations between 1994 and 2012 are also well predicted by our model. This pattern is well known to be clustered (80% of the nucleotide fixations occur in a subset of 11 years), which reflects recurrent selective sweeps in the evolution of influenza HA^{4–7,9}.

Figure 2c maps our prediction onto the strain tree. Each clade, represented by its ancestor strain, is coloured according to the maximum of the predicted frequency changes $\hat{X}_v(t+1) - X_v(t)$ over its history. We find clade expansion predominantly close to the trunk and decline far away from the trunk, which is consistent with the observed shape of the influenza tree.

Tagging clades by their point of origination (Methods), allows us to analyse correlations between fitness and geographical location. For clades originating in east and southeast Asia, we predict growth ($\hat{W}_v > 1$) in 77% of the cases, compared to 54% of the cases for clades originating elsewhere; the corresponding fractions with observed growth ($W_v > 1$) are 77% and 49%. This is consistent with the particular role of east and southeast Asia in seeding antigenic variants, which has been established previously¹⁸. Thus, our analysis captures broad spatial patterns in the evolution of influenza A/H3N2, although the underlying fitness model is geographically neutral (this point will be discussed further below).

We can quantify the statistical information gain due to our prediction by comparing distributions of predicted and posterior next-year frequencies (Methods). We find the observed frequency evolution to be more likely by a factor $> 10^{250}$ under our fitness model compared to a null model with constant frequencies (that is, zero fitness) for all strains. We emphasize that our prediction works only from one year to the next, because it cannot predict the new mutations that arise after its base year and shape the course of evolution over longer periods.

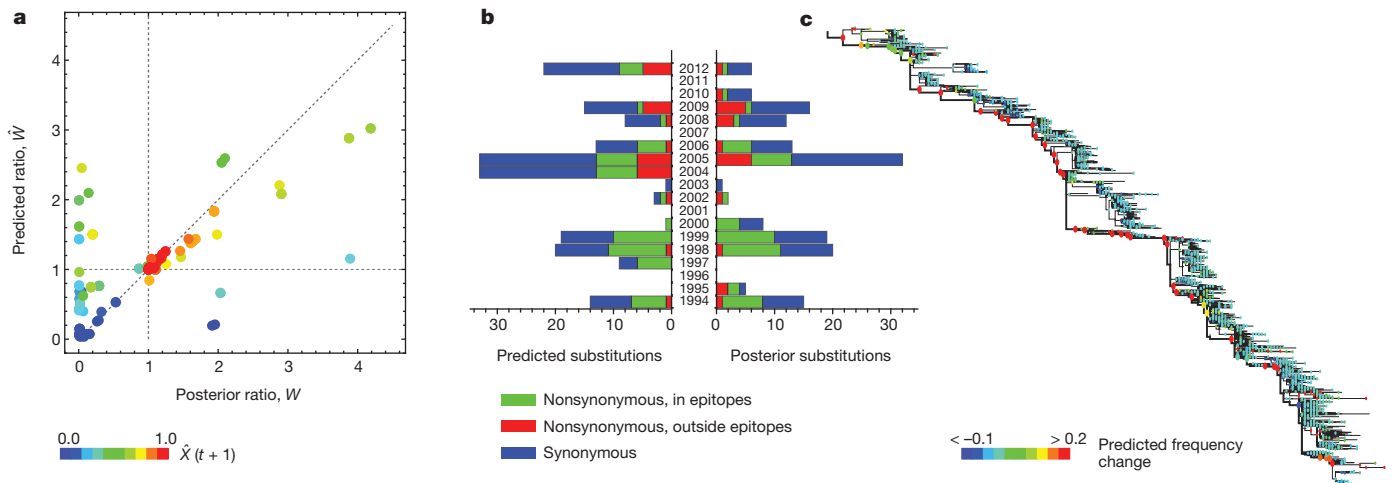


Figure 2 | Year-to-year predictions of HA evolution. **a**, Wrightian fitness: the predicted frequency ratio $\hat{W}_v = \hat{X}_v(t+1)/X_v(t)$ is plotted against the posterior ratio $W_v = X_v(t+1)/X_v(t)$ for 188 influenza HA clades with initial frequency $X_v(t) > 0.15$ observed since 1993 (error bars due to tree reconstruction and sampling are given in Extended Data Fig. 1). The predicted frequency $\hat{X}(t+1)$ is indicated by colour; clades destined for fixation are shown in red. Clade growth ($W_v > 1$) is correctly predicted in 113 of 121 cases, clade decline in 51 of

67 cases. **b**, Yearly numbers of HA nucleotide fixations: predicted numbers are compared to posterior numbers. **c**, Dynamics on the influenza strain tree: for each clade originating between 1993 and 2010, the ancestor node is coloured according to the maximum of the predicted frequency changes, $\max_t [\hat{X}_v(t+1) - X_v(t)]$. Our model correctly predicts expansion along the trunk (thick line) and loss of branches off trunk.

To test our method on a related system, we obtain clade fitness predictions for seasonal influenza A/H1N1. This lineage has re-entered the human population in 1977 and evolved in a way broadly similar to H3N2 until 2009, when the pandemic H1N1 lineage emerged. Compared to the H3N2 data set, the H1N1 strain sample^{1,25} has larger regional and seasonal biases, potentially weaker antigenic selection⁸, and larger uncertainty about the exact epitope positions²⁶ (Methods). Our predictions for H1N1 are comparable to H3N2 but somewhat more noisy, as expected from their less informative strain sample (Extended Data Fig. 2). This establishes a proof of principle for the applicability of our model to other influenza strains.

Vaccine strain selection

Our model provides a principled method to select strains for influenza vaccines. By equation (2), vaccination based on a strain v reduces the fitness of each circulating strain i proportionally to the cross-immunity amplitude $C(\mathbf{a}_i, \mathbf{a}_v)$. This causes a reduction in the total number of infections that is proportional to the average cross-immunity between the vaccine strain and the circulating strains in a given season, $C_v(t) = \sum_{i:t} x_i C(\mathbf{a}_i, \mathbf{a}_v)$ (Methods). The optimal vaccine maximizes this reduction, which defines the cross-immunity centre of mass of the circulating strains. Equation (1) predicts next-year cross-immunity amplitudes $\hat{C}_v(t+1) = \sum_{i:t} x_i \exp(f_i) C(\mathbf{a}_i, \mathbf{a}_v)$, which can be compared a posteriori with the observed amplitudes $C_v(t+1)$.

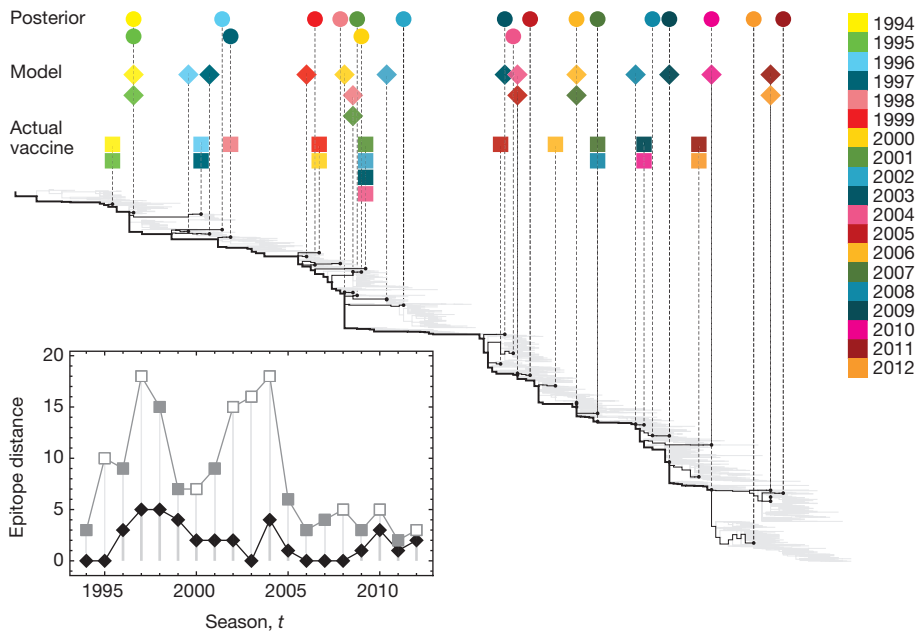
In particular, we can compare the optimal vaccine strains predicted by our model and actual vaccine strains used in the Northern Hemisphere²⁷ to the posterior centre-of-mass strains observed in the following year (the established procedure of vaccine strain selection is described in Methods). Figure 3 shows this comparison for influenza A/H3N2 in the winter seasons from 1994 to 2012. In all years, the model-selected vaccine strains have a smaller amino acid distance from the cross-immunity centre of mass of the same season than the actual vaccine strains (insert of Fig. 3). This can be explained in part by differences between our sequence-based cross-immunity measure $C(\mathbf{a}_i, \mathbf{a}_j)$ and the haemagglutination-inhibition-based antigenic distances currently used for vaccine selection. The latter are known to evolve in a more punctuated way¹², but we observe distance differences even in years when vaccine strains have been updated. These results suggest that a fitness-model-based prediction of influenza evolution can contribute to vaccine

strain selection; however, we caution against premature conclusions before our prediction scheme is carefully tested with haemagglutination inhibition data. Our model can also be used to estimate how vaccination affects the course of influenza evolution (Methods).

Mapping the adaptive process

The fitness effects underlying our predictions can be displayed in a quantitative map of influenza's adaptive history. As key quantity we use the cumulative fitness flux^{28,29}, which measures the total amount of adaptation up to a given clade; this quantity is defined in Methods and illustrated in Extended Data Fig. 3. The map of Fig. 4 shows the fitness flux for 234 influenza A/H3N2 clades on a tree between 2003 and 2008 (see Extended Data Fig. 3 for fitness flux over a longer period). It displays clades with multiple different values of fitness and fitness flux in each year. The evolution of this distribution generates a travelling fitness flux wave, which links influenza to recent theoretical models of asexual evolution^{30–34}. The advance of the wave is measured by the population mean fitness flux, which is shown as a black dashed line in Fig. 4. This quantity measures correlations between fitness and actual frequency changes of clades. It can be used to compare the predictive power of different fitness models. The best epitope-only fitness model captures about 63%, and the best model with uniform selective effects about 57% of the cumulative fitness flux given by the full model (Extended Data Table 1). An information-theoretic comparison of fitness models shows the same ranking (Methods, Extended Data Table 1). These results indicate that non-epitope changes and nonlinear selective effects have an important role in the adaptive process of influenza and its successful prediction.

The underlying mode of evolution is revealed by individual flux genealogies shown in Fig. 4. We observe that high-fitness clades seed future high-fitness clades by beneficial mutations. In particular, the clades on the trunk of the tree have consistently high predicted fitness values and are driven to fixation by multiple beneficial mutations during their expansion. At the same time, many high-fitness clades are eventually driven to loss, because they are overtaken in fitness by other competing clades. Individual beneficial alleles are lost if they arise in a low-fitness clade or if they are outcompeted by subsequent beneficial mutations in disjoint clades. These observations provide direct evidence of clonal interference in the evolution of influenza⁹ with pervasive effects of

**Figure 3 | Vaccine selection.**

Optimal vaccine strains predicted by our model (diamonds) and actual vaccine strains used in the Northern Hemisphere²⁷ (squares, listed in Supplementary Information) are compared to posterior cross-immunity centre-of-mass strains (bullets) for the winter seasons from 1994 to 2012. Model predictions are obtained by maximizing the predicted cross-immunity overlap between the vaccine strain and the circulating strains, which amounts to maximizing the predicted reduction of infections (see text and Methods). Insert: yearly epitope amino acid distances of the model-selected vaccine strain (diamonds) and the actual vaccine strain (squares, update years marked by filled squares) to the posterior cross-immunity centre-of-mass strain.

genetic linkage on individual alleles¹⁰. This mode of evolution is well known from laboratory evolution experiments with microbial and viral populations^{35,36}.

Clonal interference can explain the observed regional fitness differences between influenza A/H3N2 clades as an effect of multiple beneficial mutations coexisting in a population: individual antigenic mutations originating in east and southeast Asia have the same average effect as mutations originating elsewhere, but they occur in lineages that have accumulated more previous beneficial mutations in their recent past.

Discussion

We have developed a dynamical model that successfully predicts the year-to-year evolution of individual influenza clades, based on epitope

and non-epitope characteristics of their HA gene. Our general model is applicable whenever host–pathogen interactions—in particular, antigenic selection—generate continual adaptive evolution of a predominantly asexual population. Our results highlight the determinants of predictive power: we need sufficient information on the genotypic and phenotypic basis of antigenic and mutational-load fitness components, and model training requires a sufficiently deep and unbiased strain sample. This suggests that predictions can be improved by integrating diverse genotypic and phenotypic data, which include free-energy effects of specific mutations¹³, haemagglutination inhibition data¹², the genomics of neuraminidase⁸ and the geographical distribution of strains¹⁸. Furthermore, the prediction scheme can be extended from population frequencies to absolute growth rates and population numbers, which includes the dynamics of yearly incidence rates²³. Together, we expect an improved understanding of selective effects for specific mutations from limited strain data. This is key to evolutionary predictions for other influenza variants, including the potentially pandemic avian A/H5N1 and A/H7N9 lineages.

In a broader context, our model establishes a direct link between population genetics and epidemiology that is to be explored more comprehensively in future work. This link is the strain-specific fitness function of equation (2), which governs the dynamics of infected host individuals in an SIR framework. Strain fitness depends not only on antigenic characteristics, but also on other phenotypes encoded in genetically linked sequence. We expect that this coupling between antigenic adaptation and conservation of other functions is not limited to influenza, but is a generic feature of fast-adapting pathogens. Therefore, the epidemiology of such systems should be based on the ensemble of phenotypes linked to the adaptive process.

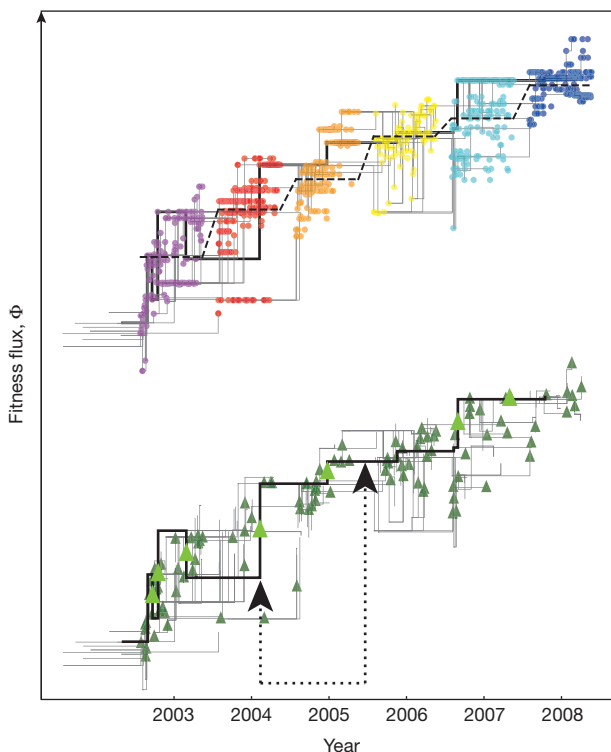


Figure 4 | Adaptation map of influenza. The fitness flux $\Phi_i(t)$, computed from the fitness model (2) and observed frequency changes, is shown for 234 clades on a tree between 2003 and 2008; see Methods for the definition and Extended Data Fig. 3a for an illustration of fitness flux. Top graph: strains within these clades are ordered by year and, within each year, by mutational distance to the last common ancestor. The mean cumulative fitness flux $\Phi(t)$ is shown as dashed line; see also Extended Data Fig. 3b. This map displays a travelling fitness flux wave. Bottom graph: the same map is shown with nonsynonymous epitope mutations marked by green triangles; these mutations are mostly beneficial^{7,9,10}. This gives evidence of clonal interference: successful clades are driven to fixation by multiple beneficial mutations (large green triangles; origination and fixation of one such clade are marked by arrows), whereas other beneficial mutations are driven to loss (small green triangles).

Beyond pathogens, this work touches upon the fundamental question of how predictable evolution is. Although there is clearly no general answer to this question, our analysis shows under what auspices limited predictions may be successful.

METHODS SUMMARY

We partition our strain sample into seasons; $t = y$ labels the period from October of year $y - 1$ to April of year y . For the prediction from season t to $t + 1$, we use maximum-likelihood HA sequence trees of all strains up to season t ; validation is based on trees for the full period. Epitope and non-epitope mutations are mapped onto the branches of these trees (Extended Data Fig. 4). Clades are defined as sets of all descendants of a given HA sequence \mathbf{a}_i ; clade frequencies are estimated by averaging over equiprobable trees.

Our fitness model has three components. The epitope fitness component $f_i^{\text{ep}} = \sigma_{\text{ep}} \sum_{j:t_j < t_i} x_j c(D_{\text{ep}}(\mathbf{a}_i, \mathbf{a}_j))$ is computed from linear amino acid distances D_{ep} between pairs of sequences in epitope codons⁷, using a nonlinear cross-immunity amplitude $c(D_{\text{ep}}) = \exp(-D_{\text{ep}}/D_0)$. The non-epitope fitness component $f_i^{\text{ne}} = -\sigma_{\text{ne}} D_{\text{ne}}(\mathbf{a}_i, \mathbf{a}_i^*)$ depends on the non-epitope amino acid distance D_{ne} between \mathbf{a}_i and the last ancestor of strain i in a previous season, \mathbf{a}_i^* . We show that these recent non-epitope mutations are under substantial negative selection⁹ (Extended Data Fig. 4). The full fitness model (equation (2)) exploits an additional feature of the population history: synonymous mutations hitchhiking in selective sweeps⁹ reinforce the inference of positive selection by a term $f_i^{\text{nl}} = \lambda D_0^{\text{ne}}(v(i), \mathbf{a}_i^*(t))$ proportional to the average neutral mutational distance of all strains in clade $v(i)$ from the last common ancestor $\mathbf{a}_i^*(t)$. The full model takes the form of equation (2), $f_i = f_i^{\text{ep}} + f_i^{\text{ne}} + f_i^{\text{nl}} + f_0$, with a constant f_0 given by the normalization conditions $\sum_i x_i = \sum_i x_i \exp(f_i) = 1$.

To quantify the quality of clade frequency predictions, we partition the strains of each season t into non-overlapping clades $v \in K(t)$ with frequencies $Y_t = (Y_v(t))_{v \in K(t)}$. We define the relative information of observed versus predicted next-year frequencies, $H(Y_{t+1} | \hat{Y}_t) = \sum_{v \in K(t)} Y_v(t+1) \log[Y_v(t+1) / \hat{Y}_v(t)]$, which determines the likelihood p that the observed strains are a sample of the predicted frequency distributions. We use this likelihood to rank fitness model variants, which include alternative epitopes and glycosylation effects, and to infer optimal parameters from strain data of 4–8 years before the base year of prediction (Extended Data Table 1 and Extended Data Fig. 6).

We define the cumulative fitness flux for each non-overlapping clade $v \in K(t)$,

$$\Phi_v(t) = F_v(t) - \bar{F}(t) + \sum_{t'=t_0}^{t-1} \sum_{v' \in K(t')} [F_{v'}(t') - \bar{F}(t')] [Y_v(t'+1) - Y_{v'}(t')]$$

where $F_v(t) = \log[\hat{Y}_v(t+1)/Y_v(t)]$ is the fitness of clade v and $\bar{F}(t) = \sum_{v \in K(t)} F_v(t) Y_v(t)$ is the mean population fitness in year t . The mean cumulative flux $\bar{\Phi}(t) = \sum_{v \in K(t-1)} \Phi_v(t-1) Y_v(t)$ is a measure of adaptation satisfying the fitness flux theorem²⁹. Deviations from this mean reflect fitness differences between clades, $\Phi_v(t) - \bar{\Phi}(t) = F_v(t) - \bar{F}(t)$.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April 2013; accepted 29 January 2014.

Published online 26 February 2014.

1. Bao, Y. *et al.* The Influenza Virus Resource at the National Center for Biotechnology Information. *J. Virol.* **82**, 596–601 (2008).
2. Wiley, D. C., Wilson, I. A. & Skehel, J. J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* **289**, 373–378 (1981).
3. Bush, R. M., Bender, C. A., Subbarao, K., Fox, N. J. & Fitch, W. M. Predicting the evolution of human influenza A. *Science* **286**, 1921–1925 (1999).
4. Plotkin, J. B., Dushoff, J. & Levin, S. A. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl Acad. Sci. USA* **99**, 6263–6268 (2002).
5. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of inter pandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903 (2006).
6. Wolf, Y. I., Viboud, C., Holmes, E. C., Koonin, E. V. & Lipman, D. J. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct* **1**, 34 (2006).
7. Shih, A. C.-C., Hsiao, T.-C., Ho, M.-S. & Li, W.-H. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc. Natl Acad. Sci. USA* **104**, 6283–6288 (2007).
8. Bhatt, S., Holmes, E. C. & Pybus, O. G. The genomic rate of molecular adaptation of the human influenza A virus. *Mol. Biol. Evol.* **28**, 2443–2451 (2011).

9. Strelkowa, N. & Lässig, M. Clonal interference in the evolution of influenza. *Genetics* **192**, 671–682 (2012).
10. Illingworth, C. J. R. & Mustonen, V. Components of selection in the evolution of the influenza virus: linkage effects beat inherent selection. *PLoS Pathog.* **8**, e1003091 (2012).
11. Meyer, A. G., Dawson, E. T. & Wilke, C. O. Cross-species comparison of site-specific evolutionary-rate variation in influenza haemagglutinin. *Phil. Trans. R. Soc. B* **368**, 20120334 (2013).
12. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of Influenza virus. *Science* **305**, 371–376 (2004).
13. Bloom, J. D. & Glassman, M. J. Inferring stabilizing mutations from protein phylogenies: application to influenza hemagglutinin. *PLOS Comput. Biol.* **5**, e1000349 (2009).
14. Wylie, C. S. & Shakhnovich, E. I. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl Acad. Sci. USA* **108**, 9916–9921 (2011).
15. Holmes, E. C. *et al.* Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* **3**, e300 (2005).
16. Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332 (2004).
17. Rambaut, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619 (2008).
18. Russell, C. A. *et al.* The global circulation of seasonal influenza A (H3N2) viruses. *Science* **320**, 340–346 (2008).
19. Gog, J. R. & Grenfell, B. T. Dynamics and selection of many-strain pathogens. *Proc. Natl Acad. Sci. USA* **99**, 17209–17214 (2002).
20. Tria, F., Lässig, M., Peliti, L. & Franz, S. A minimal stochastic model for influenza evolution. *J. Stat. Mech.* P07008 (2005).
21. Kryazhimskiy, S., Dieckmann, U., Levin, S. A. & Dushoff, J. On state-space reduction in multi-strain pathogen models, with an application to antigenic drift in influenza A. *PLOS Comput. Biol.* **3**, e159 (2007).
22. Minayev, P. & Ferguson, N. Improving the realism of deterministic multi-strain models: implications for modelling influenza A. *J. R. Soc. Interface* **6**, 509–518 (2009).
23. Rasmussen, D. A., Ratmann, O. & Koelle, K. Inference for nonlinear epidemiological models using genealogies and time series. *PLOS Comput. Biol.* **7**, e1002136 (2011).
24. Kryazhimskiy, S., Dushoff, J., Bazykin, G. A. & Plotkin, J. B. Prevalence of epistasis in the evolution of influenza A surface proteins. *PLoS Genet.* **7**, e1001301 (2011).
25. Bogner, P. *et al.* A global initiative on sharing avian flu data. *Nature* **442**, 981 (2006).
26. Huang, J.-W., Lin, W.-F. & Yang, J.-M. Antigenic sites of H1N1 influenza virus hemagglutinin revealed by natural isolates and inhibition assays. *Vaccine* **30**, 6327–6337 (2012).
27. WHO Recommendations for Influenza Vaccine Composition. Retrieved from <http://www.who.int/influenza/vaccines/vaccinerecommendations1/en>.
28. Mustonen, V. & Lässig, M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet.* **25**, 111–119 (2009).
29. Mustonen, V. & Lässig, M. Fitness flux and ubiquity of adaptive evolution. *Proc. Natl Acad. Sci. USA* **107**, 4248–4253 (2010).
30. Desai, M. M. & Fisher, D. S. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798 (2007).
31. Rouzine, I. M., Brunet, E. & Wilke, C. O. The traveling-wave approach to asexual evolution: Mullers ratchet and speed of adaptation. *Theor. Popul. Biol.* **73**, 24–46 (2008).
32. Schiffls, S., Szollosi, G. J., Mustonen, V. & Lässig, M. Emergent neutrality in adaptive asexual evolution. *Genetics* **189**, 1361–1375 (2011).
33. Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O. & Desai, M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl Acad. Sci. USA* **109**, 4950–4955 (2012).
34. Neher, R. A. & Hallatschek, O. Genealogies of rapidly adapting populations. *Proc. Natl Acad. Sci. USA* **110**, 437–442 (2013).
35. Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102–103**, 127–144 (1998).
36. Miralles, R., Gerrish, P. J., Moya, A. & Elena, S. F. Clonal interference and the evolution of RNA viruses. *Science* **285**, 1745–1747 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge discussions with B. D. Greenbaum, B. Grenfell, C. Illingworth, A. Levine, J. W. McCauley, V. Mustonen, S. Pompei and R. Rabadan. This work has been partially supported by Deutsche Forschungsgemeinschaft grant SFB 680 and by German Federal Ministry of Education and Research grant 0315893-Sybacol. Part of this work was performed at the Kavli Institute of Theoretical Physics (Santa Barbara), which has been supported by National Science Foundation grant PHY05-51164.

Author Contributions Both authors designed research, developed methods, analysed data, interpreted results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M. Lässig (lassig@thp.uni-koeln.de).

DNA interrogation by the CRISPR RNA-guided endonuclease Cas9

Samuel H. Sternberg^{1*}, Sy Redding^{2*}, Martin Jinek^{3,†}, Eric C. Greene⁴ & Jennifer A. Doudna^{1,3,5,6}

The clustered regularly interspaced short palindromic repeats (CRISPR)-associated enzyme Cas9 is an RNA-guided endonuclease that uses RNA–DNA base-pairing to target foreign DNA in bacteria. Cas9–guide RNA complexes are also effective genome engineering agents in animals and plants. Here we use single-molecule and bulk biochemical experiments to determine how Cas9–RNA interrogates DNA to find specific cleavage sites. We show that both binding and cleavage of DNA by Cas9–RNA require recognition of a short trinucleotide protospacer adjacent motif (PAM). Non-target DNA binding affinity scales with PAM density, and sequences fully complementary to the guide RNA but lacking a nearby PAM are ignored by Cas9–RNA. Competition assays provide evidence that DNA strand separation and RNA–DNA hetero-duplex formation initiate at the PAM and proceed directionally towards the distal end of the target sequence. Furthermore, PAM interactions trigger Cas9 catalytic activity. These results reveal how Cas9 uses PAM recognition to quickly identify potential target sites while scanning large DNA molecules, and to regulate scission of double-stranded DNA.

RNA-mediated adaptive immune systems in bacteria and archaea rely on CRISPRs and CRISPR-associated (Cas) proteins to provide protection from invading viruses and plasmids¹. Bacteria harbouring CRISPR–Cas loci respond to viral and plasmid challenge by integrating short fragments of the foreign nucleic acid (protospacers) into the host chromosome at one end of the CRISPR array². Transcription of the CRISPR array followed by enzymatic processing yields short CRISPR RNAs (crRNAs) that direct Cas protein-mediated cleavage of complementary target sequences within invading viral or plasmid DNA^{3–5}. In type II CRISPR–Cas systems, Cas9 functions as an RNA-guided endonuclease that uses a dual-guide RNA consisting of crRNA and *trans*-activating crRNA (tracrRNA) for target recognition and cleavage by a mechanism involving two nuclease active sites that together generate double-stranded DNA (dsDNA) breaks^{6,7}.

RNA-programmed Cas9 has proven to be a versatile tool for genome engineering in multiple cell types and organisms^{8–19}. Guided by either a natural dual-RNA complex or a chimaeric single-guide RNA⁶, Cas9 generates site-specific dsDNA breaks that are repaired either by non-homologous end joining or homologous recombination. In addition, catalytically inactive Cas9 alone or fused to transcriptional activator or repressor domains can be used to alter transcription levels at sites targeted by guide RNAs^{20–25}. Despite the ease with which this technology can be applied, the fundamental mechanism that enables Cas9–RNA to locate specific 20-base-pair (bp) DNA targets within the vast sequence space of entire genomes remains unknown.

Single-molecule visualization of Cas9

We used a single-tethered DNA curtains assay and total internal reflection fluorescence microscopy (TIRFM) to visualize the binding site distribution of single Cas9–RNA molecules on λ -DNA substrates (48,502 bp) (Fig. 1a)^{26,27}. We purified *Streptococcus pyogenes* Cas9 containing a carboxy-terminal 3 \times -Flag tag that enabled fluorescent labelling using anti-Flag-antibody-coated quantum dots (QDs)^{27,28}, and generated guide RNAs (dual crRNA–tracrRNA) bearing complementarity to six different

20-bp sites within the λ -DNA (Fig. 1b and Extended Data Table 1). Neither the 3 \times -Flag tag nor QDs inhibited DNA cleavage by Cas9–RNA, and all guide RNAs were functional (Extended Data Fig. 1). Initial experiments used a nuclease-inactive Cas9 containing D10A and H840A point mutations (dCas9) that binds but does not cleave DNA⁶. QD-tagged dCas9–RNA localized almost exclusively to the expected target site (Fig. 1c and Supplementary Video 1). Furthermore, Cas9 could be directed to any region of the DNA by redesigning the RNA guide sequence (Fig. 1d and Extended Data Fig. 2), as anticipated^{6,8,9}. Thus, DNA targeting by Cas9–RNA is faithfully recapitulated in the DNA curtains assay.

We next used *apo*-Cas9 protein to confirm that the binding observed in DNA curtains assays was due to Cas9–RNA and not *apo*-Cas9 lacking guide RNA. Notably, *apo*-Cas9 also bound DNA but exhibited no apparent sequence specificity (Fig. 1e, f). Attempts to measure the dissociation rate of DNA-bound *apo*-Cas9 were hampered by their exceedingly long lifetimes (a lower limit of at least 45 min was calculated; Fig. 1g). Biochemical experiments revealed an upper limit of ~ 25 nM for the equilibrium dissociation constant (K_d) of the *apo*-Cas9–DNA complex, compared to ~ 0.5 nM for the Cas9–RNA complex bound to a bona fide target (Extended Data Fig. 3).

To test whether DNA-bound *apo*-Cas9 could be distinguished from Cas9–RNA, we measured the lifetime of *apo*-Cas9 on DNA curtains before and after injection of crRNA–tracrRNA or heparin. *Apo*-Cas9 rapidly dissociated from nonspecific sites in the presence of either competitor (Fig. 1g), and this result was verified with bulk biochemical assays (Extended Data Fig. 3). In contrast, target-bound Cas9–RNA was unaffected by either heparin or excess crRNA–tracrRNA (Fig. 1g and Extended Data Fig. 3). These findings show that nonspecifically bound *apo*-Cas9 has properties distinct from those of Cas9–RNA complexes bound to cognate targets.

Initially, we used catalytically inactive dCas9 to avoid DNA cleavage. Notably, experiments with wild-type Cas9 also failed to reveal DNA cleavage; rather, Cas9–RNA molecules remained bound to their target sites, yielding identical results to those obtained using dCas9–RNA (Fig. 2a).

¹Department of Chemistry, University of California, Berkeley, California 94720, USA. ²Department of Chemistry, Columbia University, New York, New York 10032, USA. ³Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. ⁴Department of Biochemistry and Molecular Biophysics and Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA. ⁵Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ⁶Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [†]Present address: Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland.

*These authors contributed equally to this work.

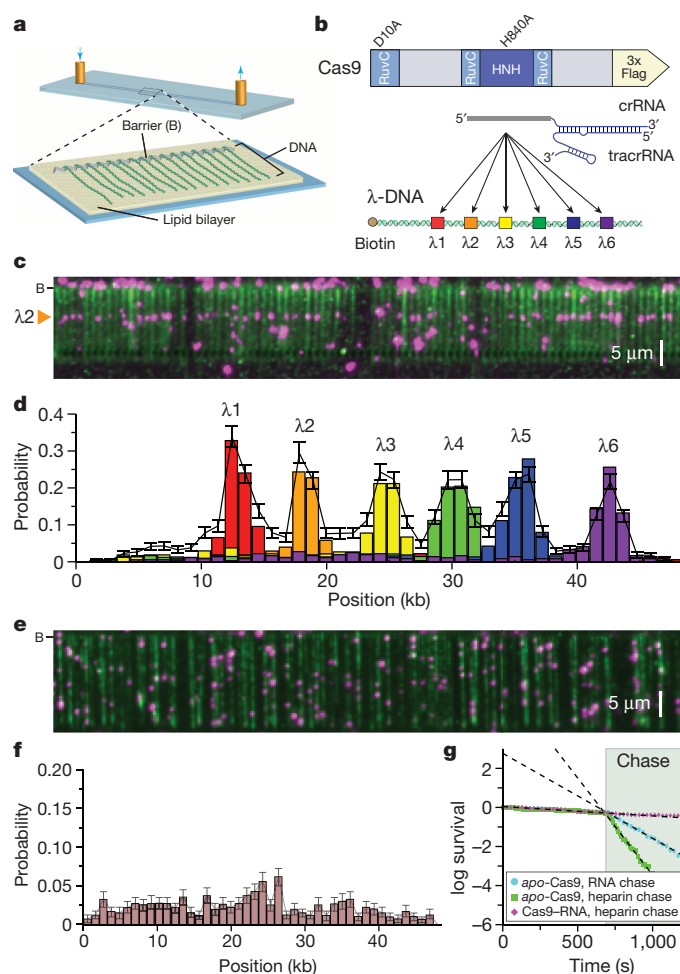


Figure 1 | DNA curtains assay for target binding by Cas9-RNA.

a, Schematic of a single-tethered DNA curtain^{26,27}. **b**, Wild-type Cas9 or dCas9 was programmed with crRNA-tracrRNA targeting one of six sites. **c**, YOYO1-stained DNA (green) bound by QD-tagged dCas9 (magenta) programmed with $\lambda 2$ guide RNA. **d**, dCas9-RNA binding distributions; error bars represent 95% confidence intervals obtained through bootstrap analysis²⁸. **e**, Image of apo-Cas9 bound to DNA curtains. **f**, Binding distribution of apo-Cas9 ($n = 467$); error bars represent 95% confidence intervals. **g**, Lifetimes of DNA-bound apo-Cas9 ($n = 205$) and Cas9-RNA ($n = 104$) after injection of $10 \mu\text{g ml}^{-1}$ heparin, and of apo-Cas9 ($n = 233$) after injection of 100 nM $\lambda 2$ crRNA-tracrRNA.

We confirmed that the imaging conditions did not inhibit Cas9-RNA cleavage activity (Extended Data Fig. 1). These results suggested that Cas9-RNA might cleave DNA but remain tightly bound to both cleavage products, which was confirmed with biochemical gel-shift assays using 72-bp duplex DNA substrates (Extended Data Fig. 4). To determine whether stable product binding would prevent Cas9-RNA from performing multiple turnover cleavage, we conducted plasmid DNA cleavage assays at varying molar ratios of Cas9-RNA and target DNA and measured the rate and yield of product formation. Surprisingly, the amount of product rapidly plateaued at a level proportional to the molar ratio of Cas9-RNA to DNA, indicating that Cas9-RNA does not follow Michaelis-Menten kinetics (Fig. 2b). Turnover also did not occur with short duplex DNA substrates and is not stimulated by either elevated temperature or an excess of free crRNA-tracrRNA (Extended Data Fig. 5).

We next used a double-tethered DNA curtain (Fig. 2c)^{29–31} to confirm that Cas9-RNA catalysed DNA cleavage in single-molecule assays. Although Cas9-RNA failed to dissociate from target sites even in the presence of $10 \mu\text{g ml}^{-1}$ heparin (Fig. 1g) or up to 0.5 M NaCl (not shown), injection of 7 M urea caused Cas9-RNA to release the downstream end containing the PAM, confirming that the DNA was cleaved at the

expected target site (Fig. 2d). These findings show that Cas9-RNA remains tightly bound to both ends of the cleaved DNA, thus acting as a single-turnover enzyme.

Cas9-RNA finds targets by three-dimensional diffusion

To determine how Cas9-RNA locates DNA targets, we visualized the target search process using double-tethered DNA curtains. Site-specific DNA-binding proteins can locate target sites by three-dimensional collisions or through facilitated diffusion processes including one-dimensional sliding, hopping, and/or intersegmental transfer³²; these mechanisms can be distinguished by single-molecule imaging^{30,31,33}. For these assays, Cas9 programmed with $\lambda 2$ guide RNA was injected into the sample chamber, buffer flow was terminated, and reactions were visualized in real time. These experiments revealed long-lived binding events at the target site and transient binding events at other sites on the DNA (Fig. 2e, f). We saw no evidence of Cas9-RNA associating with target sites by facilitated diffusion (either one-dimensional sliding and/or hopping); instead, all target association appeared to occur directly through three-dimensional collisions (Fig. 2e and Supplementary Video 2).

The shorter-lived, nonspecific binding events exhibited complex dissociation kinetics; the simplest model that describes the data is a double-exponential decay with lifetimes of ~ 3.3 and $\sim 58 \text{ s}$ (at 25 mM KCl) (Fig. 2f). These lifetimes were readily distinguished from the long lifetime of apo-Cas9 (Fig. 2f, inset). Furthermore, the experiments were conducted in the presence of a tenfold molar excess of crRNA-tracrRNA to exclude contamination from apo-Cas9. This result indicates that at least two and possibly more binding intermediates exist on the pathway towards cognate target recognition. Although nonspecific DNA binding typically involves electrostatic interactions with the phosphate backbone such that nonspecific lifetimes tend to decrease with increasing ionic strength³⁴, the lifetimes of Cas9-RNA bound at nonspecific DNA sites were not appreciably affected by salt concentration (Fig. 2f). One remarkable implication of this finding is that these Cas9-RNA non-target binding events have characteristics more commonly attributed to site-specific association^{34,35}.

To gain further insight into the nature of the target search mechanism, we measured the locations and corresponding lifetimes of all binding events (Fig. 2g). The off-target binding lifetime distributions did not vary substantially at different regions of the DNA. However, the number of observed binding events was not uniformly distributed along the substrate, indicating that some underlying feature of the λ -DNA might influence the target search. The λ -phage genome contains 5,677 PAM sites (~ 1 PAM per 8.5 bp), but it also has an unusual polar distribution of (A+T)- and (G+C)-rich sequences²⁷, which leads to an asymmetric distribution of PAMs ($5'$ -NGG- $3'$ for *S. pyogenes* Cas9) (Fig. 2g). Pearson correlation analysis revealed that the Cas9-RNA binding site distribution was positively correlated with the PAM distribution ($r = 0.59$, $P < 0.05$) (Fig. 2h). When we repeated this experiment using a guide RNA having no complementary target sites within λ -DNA (spacer 2 crRNA)⁶, we saw no change in the binding lifetimes and an even stronger correlation with the PAM distribution (Fig. 2h and Extended Data Fig. 2b, c). These results, together with the insensitivity of short-lived binding events to ionic strength, suggested that Cas9-RNA might bind specifically to PAMs and minimize interactions with non-PAM DNA while searching for potential targets.

A PAM is required for DNA interrogation

To test whether Cas9-RNA uses PAM recognition as an obligate precursor to interrogation of flanking DNA for potential guide-RNA complementarity, we used competition assays to monitor the rate of Cas9-RNA-mediated DNA cleavage (Fig. 3a, b). From these data we extracted the average amount of time that Cas9-RNA spends sampling each competitor DNA before locating and cleaving a radiolabelled substrate (Extended Data Fig. 6). In control experiments, reaction kinetics were not perturbed by the presence of an unlabelled competitor DNA lacking PAMs and bearing no sequence relationship to the crRNA,

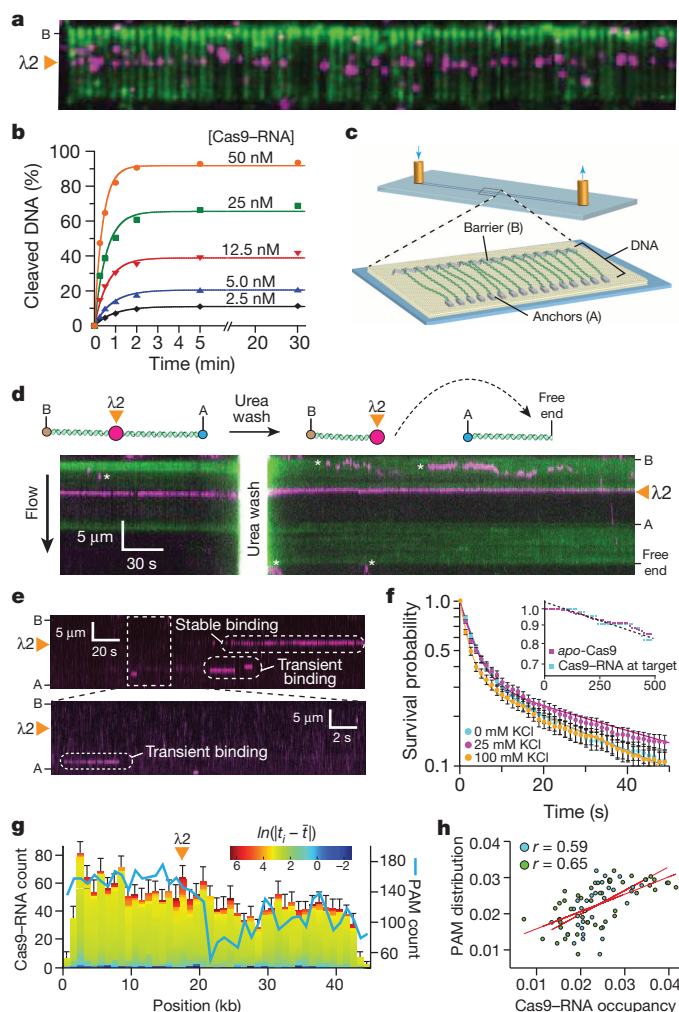


Figure 2 | Cas9-RNA remains bound to cleaved products and localizes to PAM-rich regions during the target search. **a**, Wild-type Cas9-RNA bound to DNA curtains. **b**, Cleavage yield of 25 nM plasmid DNA is proportional to [Cas9-RNA]. **c**, Schematic of a double-tethered DNA curtain^{29–31}. **d**, Liberation of the cleaved DNA with 7 M urea; asterisks denote QDs that are attached to the lipid bilayer but not bound to the DNA. **e**, Kymographs illustrating distinct binding events. **f**, Survival probabilities (mean \pm 70% confidence intervals) for non-target binding events. Solid lines represent double-exponential fits; $n = 502$ (0 mM KCl), $n = 632$ (25 mM KCl), and $n = 504$ (100 mM KCl). Inset: survival probabilities of DNA-bound *apo*-Cas9 ($n = 233$) and target DNA-bound Cas9-RNA ($n = 104$). **g**, Distribution of Cas9-RNA binding events ($n = 2,330$) and PAM sites; error bars represent 95% confidence intervals. Colour-coding reflects the binding dwell time (t_i) relative to the mean dwell time (\bar{t}). **h**, Correlation of PAM distribution and non-target Cas9-RNA binding distribution for $\lambda 2$ (blue) and spacer 2 (green) guide RNAs.

whereas a competitor containing a PAM and fully complementary target sequence substantially reduced the cleavage rate of the radiolabelled substrate (Fig. 3b).

A series of competitors were tested that bore no crRNA guide sequence complementarity (Extended Data Table 1) but contained increasing numbers of PAMs (Fig. 3c). There was a direct correspondence between the number of PAMs and the ability of a DNA competitor to interfere with target cleavage, indicating that the lifetime of Cas9-RNA on competitor DNA increased with PAM density (Fig. 3c). This result persisted over a range of competitor DNA concentrations (Extended Data Fig. 6), and the same pattern of competition was observed for DNA binding by dCas9-RNA (Extended Data Fig. 7). These results demonstrate that the residence time of Cas9-RNA on non-target DNA lacking PAMs is

negligible, and support the hypothesis that transient, non-target DNA binding events observed on the DNA curtains probably occurred at PAM sequences. Although Cas9-RNA complexes undoubtedly sample DNA lacking PAMs, these rapid binding events are neither detectable in single-molecule and bulk binding experiments (Extended Data Fig. 7), nor do they appreciably influence overall reaction kinetics in bulk biochemical assays.

We repeated the competition assay with a competitor bearing perfect complementarity to the crRNA, but with a single point mutation in the adjacent PAM (5'-TCG-3')—like similarly mutated substrates⁶, this competitor cannot be cleaved by Cas9-RNA (Extended Data Fig. 7). This competitor failed to inhibit DNA cleavage by Cas9-RNA and behaved comparably to the non-target competitor DNA lacking PAMs, despite the fact that it contained perfect complementarity to the crRNA (Fig. 3c). Together, these results demonstrate that PAM recognition is an obligate first step during target recognition by Cas9-RNA, as previously proposed⁶.

Mechanism of RNA–DNA heteroduplex formation

After PAM recognition, Cas9-RNA must destabilize the adjacent duplex and initiate strand separation to enable base pairing between the target DNA strand and the crRNA guide sequence. Because Cas9 has no energy-dependent helicase activity, the mechanism of local DNA unwinding has been enigmatic, but must rely on thermally available energy. One possibility is that PAM binding could induce a general destabilization of the duplex along the length of the entire target sequence, leading to random nucleation of the RNA–DNA heteroduplex (Fig. 3d, top). Alternatively, PAM binding may cause only local melting of the duplex, with the RNA–DNA heteroduplex nucleating at the 3' end of the target sequence next to the PAM and proceeding sequentially towards the distal 5' end of the target sequence (Fig. 3d, bottom).

To distinguish between these models, we conducted cleavage assays with a panel of DNA competitors in which the length and position of complementarity to the guide RNA was systematically varied (Extended Data Table 1). These competitors were designed to distinguish between the random nucleation and sequential unwinding models for heteroduplex formation based on the predicted patterns of cleavage inhibition for each model (Fig. 3d). The ability of a competitor DNA to inhibit substrate cleavage by Cas9-RNA increased as the extent of complementarity originating at the 3' end of the target sequence adjacent to the PAM increased (Fig. 3e). Inhibition increased markedly when 12 or more base pairs were complementary to the crRNA guide sequence, which agrees with the requirement for an 8–12-nucleotide seed sequence for the Cas9-RNA DNA cleavage reaction^{6,36}. Notably, although competitors containing mismatches to the crRNA at the 5' end of the target sequence competed effectively for Cas9-RNA binding, competitors containing mismatches to the crRNA at the extreme 3' end immediately adjacent to the PAM were completely inert to binding (Fig. 3e). This was true even with a 2-bp mismatch followed by 18 bp of contiguous sequence complementarity to the crRNA. Therefore, when mismatches to the crRNA are encountered within the first two nucleotides of the target sequence, Cas9-RNA loses the ability to interrogate and recognize the remainder of the DNA. The pattern of inhibition observed with the different competitor DNAs indicates that sequence homology adjacent to the PAM is necessary to initiate target duplex unwinding until the reaction has proceeded sufficiently far (~ 12 bp, approximately one turn of an A-form RNA–DNA helix), such that the energy necessary for further propagation of the RNA–DNA heteroduplex falls below the energy needed for the reverse reaction. These findings indicate that formation of the RNA–DNA heteroduplex initiates at the PAM and proceeds through the target sequence by a sequential, step-wise unwinding mechanism consistent with a Brownian ratchet³⁷.

As a further test of this model, we used a DNA competitor that contained mismatches to the crRNA at positions 1–2 but was itself mismatched at the same two positions, forming a small bubble in the duplex. Despite the absence of sequence complementarity to the crRNA within

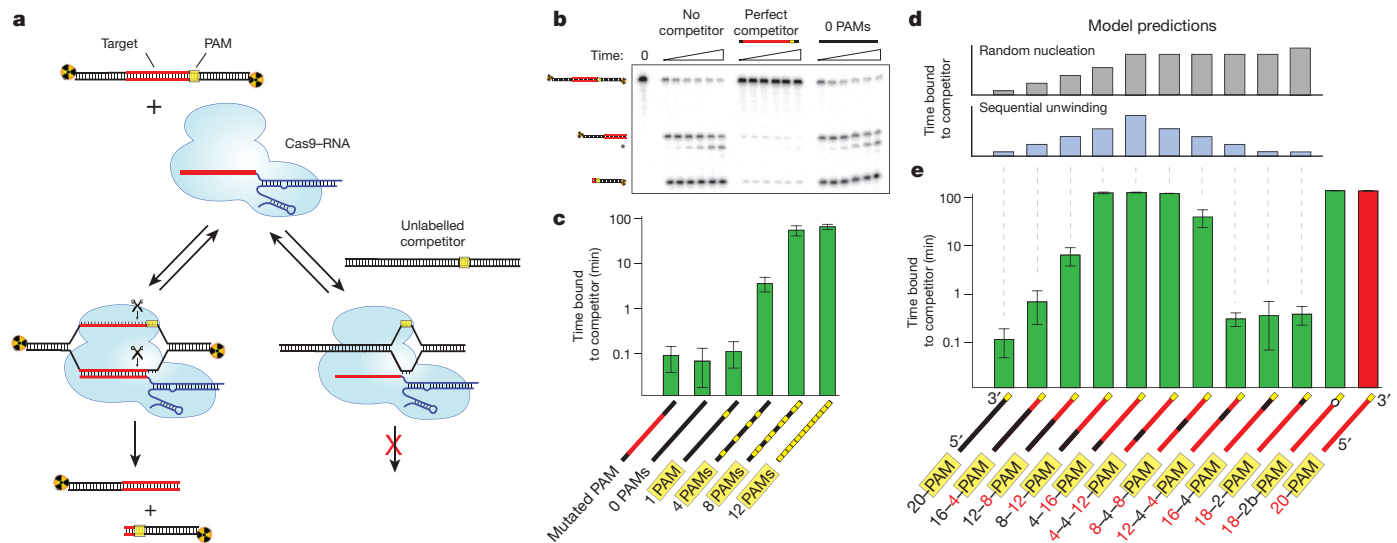


Figure 3 | Cas9-RNA searches for PAMs and unwinds dsDNA in a directional manner. **a**, Schematic of the competition cleavage assay. **b**, Cleavage assay with and without competitor DNAs. **c**, Quantification of competition data (mean \pm s.d.). Competitor cartoon representations show PAMs (yellow) and regions complementary to the crRNA (red). **d**, Predicted

data trends for the random nucleation or sequential unwinding models aligned with the corresponding data in **e**. **e**, Competition assays using substrates with variable degrees of crRNA complementarity, shown as in **c**. Numeric descriptions of the competitor DNAs indicate the regions of complementarity (red) or mismatches (black) to the crRNA sequence.

the DNA bubble, this substrate was a robust competitor and bound Cas9-RNA with an affinity nearly indistinguishable from that of a perfect substrate (Fig. 3e and Extended Data Fig. 8). Notably, this DNA could also be cleaved with near-wild-type rates (Extended Data Fig. 8). We speculate that the presence of the DNA bubble allowed Cas9-RNA to bypass the mismatches and reinitiate nucleation of the RNA-DNA

heteroduplex downstream of the bubble, thereby propagating strand separation through the remainder of the target.

The PAM triggers Cas9 nuclease activity

The results presented above indicate that PAM recognition has a central role in target recognition, and that introduction of a small bubble in

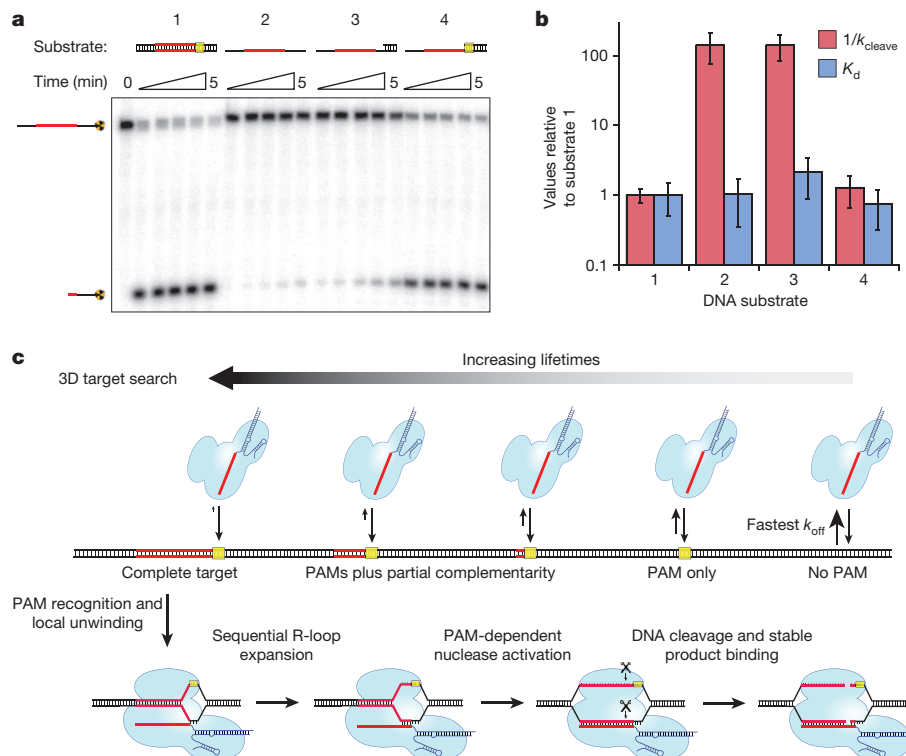


Figure 4 | PAM recognition regulates Cas9 nuclease activity. **a**, Cleavage assay with single-stranded, double-stranded and partially double-stranded substrates. **b**, Relative affinities and cleavage rates (mean \pm s.d.). **c**, Model for target search, recognition and cleavage by Cas9-RNA. The search initiates through random three-dimensional collisions. Cas9-RNA rapidly dissociates from non-PAM DNA, but binds PAMs for longer times and samples adjacent

DNA for guide RNA complementarity, giving rise to a heterogeneous population of intermediates. At correct targets, Cas9-RNA initiates formation of an RNA-DNA heteroduplex, and R-loop expansion propagates via sequential unwinding. The DNA is cleaved, and Cas9-RNA remains bound to the cleaved products.

the DNA target eliminates the need for RNA–DNA heteroduplex formation immediately adjacent to the PAM. One might expect PAM recognition to be dispensable for Cas9–RNA-mediated recognition and cleavage of a single-stranded DNA (ssDNA) target. Surprisingly, however, a ssDNA substrate was cleaved more than two orders of magnitude slower than a dsDNA substrate (Fig. 4a, b), despite the fact that dCas9–RNA bound both dsDNA and ssDNA substrates with similar affinities (Fig. 4b and Extended Data Fig. 9).

Importantly, Cas9–RNA recognizes the 5′-NGG-3′ PAM on the non-target DNA strand⁶, so the ssDNA substrates did not contain a PAM but rather the complement to the PAM sequence. We proposed that the absence of the PAM on the ssDNA might explain why an otherwise fully complementary target is resistant to cleavage. To test this possibility, we prepared hybrid substrates with varying lengths of dsDNA at the 3′ flanking sequence (Fig. 4a). Cleavage assays revealed that the ssDNA target strand could be activated for cleavage in the presence of flanking dsDNA that extended across the PAM sequence, but that this activating effect was lost when the dsDNA was truncated immediately before the PAM (Fig. 4a, b and Extended Data Fig. 9). Gel-shift experiments confirmed that these results were not a consequence of discrimination at the level of binding (Fig. 4b). Rather, the presence of the 5′-NGG-3′ PAM on the non-target strand was critical for some step of the reaction that occurred after binding. These data suggest that the PAM acts as an allosteric regulator of Cas9–RNA nuclease activity.

Discussion

Our results suggest a model for target binding and cleavage by Cas9–RNA involving an unanticipated level of importance for PAM sequences at each stage of the reaction (Fig. 4c). Although details may differ, we propose that PAM interactions may function similarly for other CRISPR RNA-guided surveillance complexes^{38–44}. The Cas9–RNA target search begins with random collisions with DNA. However, rather than sampling all DNA equivalently, Cas9–RNA accelerates the search by rapidly dissociating from non-PAM sites, thereby reducing the amount of time spent at off-targets. Only upon binding to a PAM site does Cas9–RNA interrogate the flanking DNA for guide RNA complementarity, as was previously proposed for Cas9 (ref. 6) and a distinct CRISPR RNA-guided complex (Cascade)⁴². A requirement for initial PAM recognition also eliminates the potential for self-targeting, as perfectly matching targets within the bacterial CRISPR locus are not flanked by PAMs. Our results indicate that PAM recognition coincides with initial destabilization of the adjacent sequence, as demonstrated by experiments using a bubble-containing DNA substrate, followed by sequential extension of the RNA–DNA heteroduplex. This mechanism explains the emergence of seed sequences, because mismatches encountered early in a directional melting-in process would prematurely abort target interrogation. Moreover, the complex dissociation kinetics observed on non-target λ-DNA would arise from heterogeneity in the potential target sites as Cas9–RNA probes sequences adjacent to PAMs for guide RNA complementarity. Binding to a correct target then leads to activation of both nuclease domains. This step also requires PAM recognition, providing an unanticipated level of PAM-dependent regulation that may ensure further protection against self-cleavage of the CRISPR locus. Notably, Cas9–RNA does not dissociate from the cleaved DNA except under extremely harsh conditions, indicating that Cas9–RNA may remain bound to the cleaved site *in vivo*⁴ and require other cellular factors to promote recycling. Finally, our data indicate that efforts to minimize off-target effects during genome engineering using Cas9–RNA complexes need only consider off-targets adjacent to a PAM, because potential targets lacking a PAM are unlikely to be interrogated^{36,45–47}.

METHODS SUMMARY

Cas9 from *S. pyogenes* containing a C-terminal 3×-Flag tag was purified as described⁶. crRNAs and tracrRNAs were either transcribed *in vitro* with T7 polymerase or made synthetically. Cas9–RNA complexes for single-molecule experiments were reconstituted by incubating Cas9 and a 10× molar excess of crRNA–tracrRNA in

reaction buffer (20 mM Tris–HCl pH 7.5, 100 mM KCl, 5 mM MgCl₂, 5% glycerol, 1 mM dithiothreitol (DTT)) for ~10 min at 37 °C, before injecting 1–2 nm into the flow cell. Single-tethered DNA curtains were prepared and TIRFM conducted as described^{26,27}. dCas9–RNA positions were determined by fitting a two-dimensional Gaussian to individual molecules. Double-tethered DNA curtains were prepared as described^{31,33}, and position and lifetime measurements were determined from kymographs generated for each DNA molecule. Duplex DNA substrates used in biochemical experiments were prepared from synthetic oligonucleotides (Integrated DNA Technologies), and their sequences can be found in Extended Data Table 1. Bulk competition cleavage assays were conducted at room temperature in reaction buffer and contained ~1 nM radiolabelled λ1 target DNA, 10 nM Cas9–RNA complex and 500 nM competitor DNA. Aliquots were removed at each time point and quenched by the addition of gel loading buffer supplemented with 50 mM EDTA. Reaction products were resolved by 10% denaturing polyacrylamide gel electrophoresis, visualized by phosphorimaging, and quantified using ImageQuant (GE Healthcare). Competition experiments were analysed to determine the survival probability of the target DNA. After determining the conditional survival probability for each time course, the change in survival probability relative to a reaction in the absence of competitor was calculated. Integration of the resulting curve over the total reaction time yielded the mean time spent on competitor DNA for each experiment.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 September; accepted 23 December 2013.

Published online 29 January 2014.

- Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
- Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
- Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
- Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
- Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
- Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
- Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnol.* **31**, 227–229 (2013).
- Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
- Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 System. *Cell Rep.* **4**, 220–228 (2013).
- Gratz, S. J. *et al.* Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029–1035 (2013).
- Friedland, A. E. *et al.* Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature Methods* **10**, 741–743 (2013).
- Shan, Q. *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnol.* **31**, 686–688 (2013).
- Li, J.-F. *et al.* Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature Biotechnol.* **31**, 688–691 (2013).
- Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G. & Kamoun, S. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnol.* **31**, 691–693 (2013).
- Xie, K. & Yang, Y. RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* **6**, 1975–1983 (2013).
- Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
- Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
- Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
- Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10**, 977–979 (2013).
- Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10**, 973–976 (2013).

25. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnol.* **31**, 833–838 (2013).
26. Fazio, T., Visnapuu, M.-L., Wind, S. & Greene, E. C. DNA curtains and nanoscale curtain rods: high-throughput tools for single molecule imaging. *Langmuir* **24**, 10524–10531 (2008).
27. Visnapuu, M.-L. & Greene, E. C. Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nature Struct. Mol. Biol.* **16**, 1056–1062 (2009).
28. Finkelstein, I. J., Visnapuu, M.-L. & Greene, E. C. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature* **468**, 983–987 (2010).
29. Gorman, J., Fazio, T., Wang, F., Wind, S. & Greene, E. C. Nanofabricated racks of aligned and anchored DNA substrates for single-molecule imaging. *Langmuir* **26**, 1372–1379 (2010).
30. Gorman, J., Plys, A. J., Visnapuu, M.-L., Alani, E. & Greene, E. C. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nature Struct. Mol. Biol.* **17**, 932–938 (2010).
31. Wang, F. *et al.* The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion. *Nature Struct. Mol. Biol.* **20**, 174–181 (2013).
32. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
33. Gorman, J. *et al.* Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl Acad. Sci. USA* **109**, E3074–E3083 (2012).
34. von Hippel, P. H. & Berg, O. G. On the specificity of DNA-protein interactions. *Proc. Natl Acad. Sci. USA* **83**, 1608–1612 (1986).
35. Rohs, R. *et al.* Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
36. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnol.* **31**, 233–239 (2013).
37. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).
38. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
39. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
40. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108**, 10098–10103 (2011).
41. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl Acad. Sci. USA* **108**, 10092–10097 (2011).
42. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* **46**, 606–615 (2012).
43. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **110**, 15644–15649 (2013).
44. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116–1121 (2013).
45. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnol.* **31**, 839–843 (2013).
46. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31**, 827–832 (2013).
47. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnol.* **31**, 822–826 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Bhat, A. Smith and K. Zhou for technical assistance, and members of the Doudna and Greene laboratories and J. Cate for discussions and critical reading of the manuscript. S.H.S. acknowledges support from the National Science Foundation and National Defense Science & Engineering Graduate Research Fellowship programs. Funding was provided by the National Institutes of Health (GM074739 to E.C.G.) and the National Science Foundation (MCB-1154511 to E.C.G. and MCB-1244557 to J.A.D.). M.J. was a Research Specialist, E.C.G. is an Early Career Scientist, and J.A.D. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions S.H.S. generated RNAs, conducted biochemical and single-molecule experiments, and assisted with single-molecule data analysis. S.R. conducted single-molecule experiments and data analysis, and assisted with the design and analysis of biochemical assays. M.J. cloned and purified Cas9, and assisted with the design and interpretation of initial single-molecule experiments. S.H.S., S.R., M.J., E.C.G. and J.A.D. discussed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to J.A.D. (doudna@berkeley.edu) or E.C.G. (ecg2108@columbia.edu).

Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1

Joanne L. Parker¹ & Simon Newstead^{1,2}

The NRT1/PTR family of proton-coupled transporters are responsible for nitrogen assimilation in eukaryotes and bacteria through the uptake of peptides. However, in most plant species members of this family have evolved to transport nitrate as well as additional secondary metabolites and hormones. In response to falling nitrate levels, NRT1.1 is phosphorylated on an intracellular threonine that switches the transporter from a low-affinity to high-affinity state. Here we present both the apo and nitrate-bound crystal structures of *Arabidopsis thaliana* NRT1.1, which together with *in vitro* binding and transport data identify a key role for His 356 in nitrate binding. Our data support a model whereby phosphorylation increases structural flexibility and in turn the rate of transport. Comparison with peptide transporters further reveals how the NRT1/PTR family has evolved to recognize diverse nitrogenous ligands, while maintaining elements of a conserved coupling mechanism within this superfamily of nutrient transporters.

Nitrogen is a key element in biology, required for the synthesis of amino and nucleic acids and is a fundamental nutrient for cellular metabolism. The PTR or peptide transporter family, also known as the POT, or proton dependent oligopeptide transporter family, has an important role in nitrogen assimilation in bacteria, fungi and mammals through the uptake of short peptides from the environment^{1,2}. In contrast, in the majority of plant species nitrogen is largely obtained through the uptake of nitrate (NO_3^-) from the soil³. This is achieved through the actions of specific nitrate transporters in the plasma membrane of root cells⁴ and is controlled through the NRT1 and NRT2 gene families that encode for low ($K_M \approx \text{mM}$) and high ($K_M \approx \mu\text{M}$) affinity transporters, respectively^{5,6}. Intriguingly, the NRT1 family of nitrate transporters phylogenetically belongs to the PTR family, and may have evolved from an ancestral peptide transport protein⁷.

In plants the NRT1/PTR family, recently renamed the NPF (NRT1/PTR family)⁸, has functionally diverged with individual members recognizing peptides, glucosinolate defence compounds, plant hormones or nitrate^{9–12}. The NRT1/PTR family belongs to the major facilitator superfamily (MFS) of secondary active transporters^{13,14} that use the proton electrochemical gradient to drive substrate uptake into the cell^{15–17}. MFS transporters operate through an alternating access mechanism, wherein a centrally located binding site is re-orientated to either side of the membrane to uptake and release substrate^{18–20}. Recent crystal structures of bacterial members of the NRT1/PTR family have identified pairs of salt bridges that orchestrate the orientation of the transporter and revealed proton binding residues that are conserved across the family from prokaryotes to eukaryotes^{21,22}. An important question therefore is how the binding site in this transporter family has evolved to recognize and transport diverse molecules while retaining elements of the same coupling mechanism.

The *Arabidopsis thaliana* nitrate transporter, NRT1.1 (NPF6.3), is the founding member⁷ of the NRT1/PTR family of transporters and shares significant sequence identity to mammalian and bacterial PTR peptide transporters (Extended Data Fig. 1). Interestingly, NRT1.1 has two K_M values for nitrate²³; in conditions of high nitrate availability ($>1 \text{ mM}$) NRT1.1 behaves as a low-affinity transporter ($K_M \approx 4 \text{ mM}$). However, when nitrate levels fall below 1 mM , NRT1.1 is switched into a high-affinity mode ($K_M \approx 40 \mu\text{M}$). This switch occurs following

phosphorylation of an intracellular threonine, Thr 101, by the kinase CIPK23²⁴. This regulatory mechanism allows for the rapid adaption to changing nitrate levels before the dedicated high affinity NRT2 transporter family is expressed²⁵. These observations suggest a complex interplay between transporter activity and post-translational modification at the molecular level that currently has no structural or biochemical basis for explanation.

Structural basis of nitrate recognition in NRT1.1

To study the binding of nitrate to NRT1.1 we developed a microscale thermophoresis assay^{26,27} (Extended Data Fig. 2a) and determined the K_D for nitrate to be $1 \pm 0.15 \text{ mM}$ (Fig. 1a and Extended Data Fig. 2b). A wide range of nitrate concentrations were tested ($1 \mu\text{M}$ – 50 mM) and we did not observe a high affinity state. We further tested a range of different ligands and show that NRT1.1 specifically recognizes both nitrate and chlorate, a herbicide for which the NRT1.1 protein was originally named (chlorate resistance 1)⁷ but does not recognize nitrite, alanine, sulphate, phosphate or the di-peptide Ala-Ala (Extended Data Fig. 2c). A key question surrounding the function of NRT1.1 is the switch between high and low affinity states. Using the phosphomimetic variant Thr101Asp, we investigated the effect on nitrate binding. The K_D for nitrate binding to the Thr101Asp variant was the same as for wild type ($K_D = 1 \pm 0.12 \text{ mM}$) (Fig. 1a), indicating that phosphorylation is unlikely to alter the nitrate-binding site directly. To follow up this finding and identify the nitrate-binding site we crystallized NRT1.1 in the presence and absence of nitrate. The crystal structures were determined to 3.7 \AA (Extended Data Table 1).

The apo structure of NRT1.1 contains 12 transmembrane (TM)-spanning alpha helices, consisting of amino-(TM1–TM6) and carboxy-terminal (TM7–TM12) bundles forming the canonical MFS fold (Fig. 1b). Separating these two bundles is a previously uncharacterized intracellular domain consisting of 84 amino acids. This domain is predominantly alpha helical and extends outwards from the transporter domain at an approximately 90° angle, which we have termed the lateral helix (Fig. 1c). At the distal end of this helix are three conserved positively charged residues that may help to stabilize this domain on the intracellular side of the membrane. It is unlikely that this domain is a crystallographic artefact as it does not make any crystal contacts and is sufficiently

¹Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. ²Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, UK.

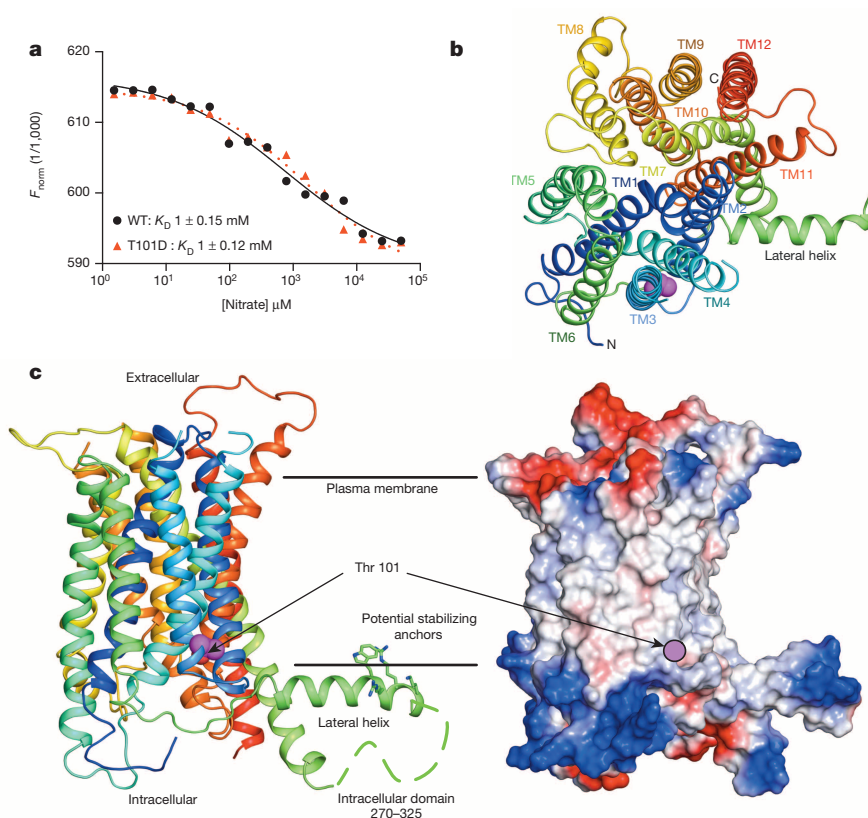


Figure 1 | Structural and biophysical characterization of NRT1.1. **a**, Binding isotherms for nitrate to both wild type (black circles) and Thr101Asp (red triangles) reveals no significant change in the K_D . Data shown is representative of three independent experiments and the error is calculated as standard deviation. **b**, Cartoon representation of the crystal structure of NRT1.1 viewed from the extracellular side of the plasma membrane. Transmembrane helices (TM) 1–12 have been coloured from blue at the amino terminus to red at the carboxy terminus. Thr 101 is shown as magenta spheres. **c**, NRT1.1 viewed in the plane of the membrane represented as both cartoon and electrostatic surface, positively charged and aromatic residues at the end of the lateral helix are shown in sticks and may help to anchor this helix in the interfacial region of the plasma membrane. Residues 270–325 of the intracellular domain could not be modelled due to insufficient electron density.

mobile that we cannot model residues 270–325. NRT1.1 crystallizes as a dimer in the asymmetric unit cell, despite being monomeric in solution (Extended Data Fig. 3). The dimer interface forms between TM3 and TM6 of monomer A packing against the equivalent helices in monomer B and stabilized by hydrophobic interactions (Extended Data Fig. 3e). The buried surface area between the two monomers is approximately $2,136 \text{ \AA}^2$ and may be physiologically significant as both monomers are in the same orientation. Both monomers A and B adopt the same inward open conformation (with a root mean squared deviation (r.m.s.d.) between monomers A and B of $\sim 0.69 \text{ \AA}$ for $472 C_\alpha$ atoms) with a clearly defined cavity extending outward from the middle of the protein towards the intracellular side of the membrane. Surprisingly, the cavity is much larger than expected given the size of nitrate, being approximately $21 \times 18 \times 17 \text{ \AA}$ in diameter (Fig. 2a). The extracellular side of the binding site is tightly sealed through the packing together of TM1–TM2 against TM7–TM8. These helices form the extracellular gate in MFS transporters, which control access to the binding site from the outside of the cell^{19,28,29}. Interestingly, there exist no salt bridge interactions between these helices, unlike the prokaryotic NRT1/PTR homologues^{14,21,22,30}; instead the extracellular gate is stabilized through extensive hydrophobic interactions (Extended Data Fig. 4).

The nitrate-bound structure was unchanged from the apo state with an r.m.s.d. of approximately 0.44 \AA for $472 C_\alpha$ atoms. However, we observed a significant ($>3\sigma$) $mFo - DFc$ difference electron density peak within the central cavity, which we observed in crystals grown in the presence of nitrate. We therefore modelled nitrate into this density (Fig. 2a, b and Extended Data Fig. 5). The nitrate molecule sits approximately 2.8 \AA away from His 356 on TM7; given the pH of the crystallization condition was 4.5 His 356 is likely to be protonated and forming an electrostatic interaction with nitrate. An additional interaction via a hydrogen bond occurs with Thr 360 on TM7, which sits approximately 2.9 \AA away at the apex of the binding site. The binding of nitrate through an electrostatic interaction is similar, but not identical, to the recent crystal structure of NarU, a member of the biochemically distinct nitrate

nitrite porter (NNP) family, where two conserved arginine residues act to coordinate nitrate through hydrogen bonds (Extended Data Fig. 6)³¹. The other possible sites of interaction for nitrate therefore could be either Arg 45 (TM1) and/or Lys 164 (TM4). We mutated all three of these positively charged residues to alanine; however, mutation of only His 356 resulted in complete loss of nitrate binding, indicating an essential role of His 356 in NRT1.1 (Fig. 2c).

Substrate specificity in the NRT1/PTR family

As discussed previously, members of the NRT1/PTR family can transport peptides and in the case of plant members, nitrate, hormones and metabolites. A key question we wished to address was how the binding site of NRT1.1 differed to that of the peptide transporters. Comparison with the binding site from the bacterial peptide transporter from *Shewanella oneidensis*, PepT_{So} , reveals that the two binding sites are strikingly similar (Fig. 2d). The previously identified ExxER motif on TM1, which has an important role in coupling proton binding to peptide transport²² is present in the same position, as are Lys 164 (TM4) and Glu 476 (TM10). As predicted for their bacterial counterparts^{21,22} Lys 164 and Glu 476 are likely to interact via a salt bridge in the outward facing conformation and may form an important link between the N and C-terminal bundles during transport (Extended Data Fig. 7).

Our comparison indicates that the mechanism required for coupling proton movement to structural changes during transport is conserved regardless of substrate specificity. It is known the stoichiometry of NRT1.1 and mammalian peptide transporters is at least 2:1 proton:solute, as transport is electrogenic^{7,32}. It is therefore conceivable the ExxER motif couples one proton, leaving His 356 on TM7 to bind another proton and nitrate. In peptide transporters the equivalent region on TM7 has been shown to couple proton binding to peptide recognition²² through an acidic residue, Asp 316 in PepT_{So} . One major difference between the binding sites is the absence of an obvious salt bridge connecting the extracellular gate helices TM1 and TM7. Indeed in PepT_{So} the arginine of this salt bridge occupies the same position in the protein as the

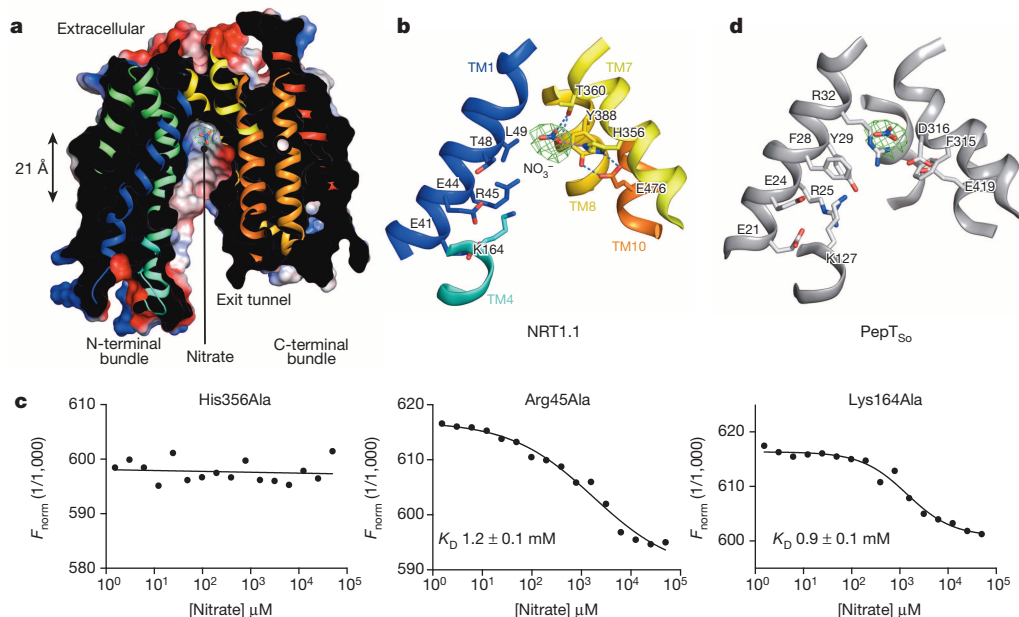


Figure 2 | Crystal structure of nitrate bound NRT1.1. **a**, Section through the protein volume showing the binding site in the plane of the membrane, with the $mF_o - DF_c$ difference electron map contoured at 3σ (green) identifying the position of nitrate. **b**, Zoomed in view of the binding site. Shown in sticks are the conserved NRT1/PTR residues and His 356 that can be seen coordinating nitrate within the binding site, supported by Thr 360. Dashed lines indicate a potential interaction network between His 356, Tyr 388 and Glu 476 in the

C-terminal bundle. **c**, Binding isotherms showing the effect of mutating positively charged residues in the binding site on the affinity of NRT1.1 for nitrate. Data shown is representative of three independent experiments and the error is calculated as standard deviation. **d**, Similarity between the binding sites of the peptide transporter PepT_{so} and NRT1.1. The residues important for peptide transport are shown in sticks. The relative position of nitrate from the NRT1.1 structure is superimposed (green mesh).

nitrate (Fig. 2b, d). Another significant difference is the replacement of the two conserved aromatic residues on TM1 (Phe 28, Tyr 29), which are responsible for peptide specificity²¹ with residues containing smaller side chains helping to create space for nitrate to bind. Given that dipeptides are coordinated horizontally between the N- and C-terminal bundles^{22,30}, it is conceivable that the increased width of the cavity in NRT1.1 can no longer coordinate the amino and carboxy termini simultaneously. Coupled with the absence of the salt bridge and tyrosines, which help position the peptide and coordinate movement of TM1 and TM7, these alterations could explain why NRT1.1 cannot recognize or transport peptides.

Functional role of phosphorylation at Thr 101

Our structures show that Thr 101 is situated at the bottom of TM3 and points towards a hydrophobic pocket constructed from residues in TM2 and TM4 (Fig. 3a). Given the position of Thr 101 it is highly likely that phosphorylation would cause localized structural distortion in the packing of TM3 with TM1 and TM4 on the intracellular side of the protein. Consistent with this hypothesis we discovered that the Thr101Asp variant of NRT1.1 was significantly less stable compared to wild-type protein (9 °C lower melting temperature) (Fig. 3b). To investigate the effect of the Thr101Asp substitution on nitrate uptake we reconstituted NRT1.1 into liposomes and monitored the relative uptake of nitrate as a function of proton movement using a pH-sensitive fluorophore (Extended Data Fig. 8). As NRT1.1 is a proton-coupled nitrate transporter this assay provides a reliable readout for nitrate transport. We discovered that compared to the wild-type protein, the Thr101Asp variant showed increased uptake of approximately 2.8-fold, whereas the His356Ala variant showed no detectable transport, consistent with an essential role in nitrate recognition (Fig. 3c).

Given the high sequence conservation of the N-terminal bundle within the NRT1/PTR family (Extended Data Figs 1 and 9a), we mutated the equivalent threonine in PepT_{so} (Thr 87) to investigate any general effects disrupting the packing between these helices may have on this family. Interestingly the Thr87Asp variant also resulted in an increase in the

rate of peptide transport, whereas other variants Thr87Ala/Ser/Lys showed wild-type-like rates (Extended Data Fig. 9b–d), showing that this region is sensitive to the addition of a negative charge. Comparable to what we observed for NRT1.1, the Thr87Asp variant was also less stable than wild-type protein by 4 °C (Extended Data Fig. 9e). However, a Thr87Arg variant was inactive, indicating that gross structural distortion in this region cannot be tolerated. This shows that disrupting the helix packing

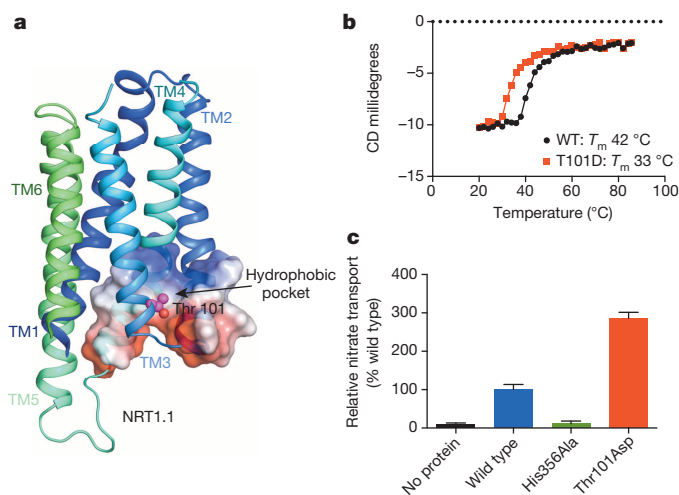


Figure 3 | Proposed mechanism for the effect of phosphorylation at Thr 101 on nitrate transport. **a**, Thr 101 (represented as magenta spheres) in NRT1.1 is situated in a hydrophobic cavity formed at the intracellular ends of TM2 and TM4. **b**, Thermal stability as assessed by the loss of alpha helical secondary structure using circular dichroism shows a reduction in the Thr101Asp variant when compared to wild-type protein. **c**, The Thr101Asp variant of NRT1.1 shows a significant increase in transport when compared to wild-type protein in a reconstituted assay system, whereas the His356Ala mutant shows no activity. Data represent the mean value from eight experiments and error bars are standard deviation.

between TMs 1, 2, 3 and 4 in the N-terminal bundle of the NRT1/PTR family can markedly affect transport rates. Similar studies have shown that altering the structural flexibility within the MFS fold can affect the rate-limiting step of transport^{33,34}. We propose that a similar mechanism may occur in NRT1.1 to switch the protein between a high and low K_M state. Phosphorylation causes a localized disruption of the N-terminal bundle helices that leads to an increased flexibility of the protein. This in turn leads to an increase in the transport rate, that would result in the lower K_M observed in the previous study²⁴.

Model for proton-coupled nitrate transport via NRT1.1

Taken together, a working model for proton-coupled nitrate transport emerges from our study (Fig. 4). The protonation of His 356 is clearly a key component of the transport mechanism. Our model predicts that protonation must occur before nitrate binding and may be conditional upon the presence of nitrate to stabilize the charged state of His 356. Following closure of the extracellular gate and adoption of the occluded state, the release of both proton and nitrate from His 356 must be coupled to the opening of the intracellular gate. A possible mechanism in achieving this is that adoption of the occluded state brings His 356 into close proximity to Glu 476, causing the disruption of the intracellular gate salt bridge (Extended Data Fig. 7). The interaction between His 356 and Glu 476 may be facilitated by Tyr 388 (TM8), which can be seen making a hydrogen bond to Glu 476 in the crystal structures. The binding of additional protons, potentially to the ExxER motif, might also affect the strength of the salt bridge between Lys 164 and Glu 476, facilitating opening of the intracellular gate and release of bound nitrate and protons into the interior of the cell. Our data suggest that in NRT1.1 the effect of phosphorylation is to cause a localized disruption in the N-terminal bundle that in turn may affect the position of the ExxER motif with respect to Lys 164. The increase in transport rate that occurs is possibly the result of altering the efficiency in the formation and breakage of the Lys 164–Glu 476 salt bridge and release of nitrate from

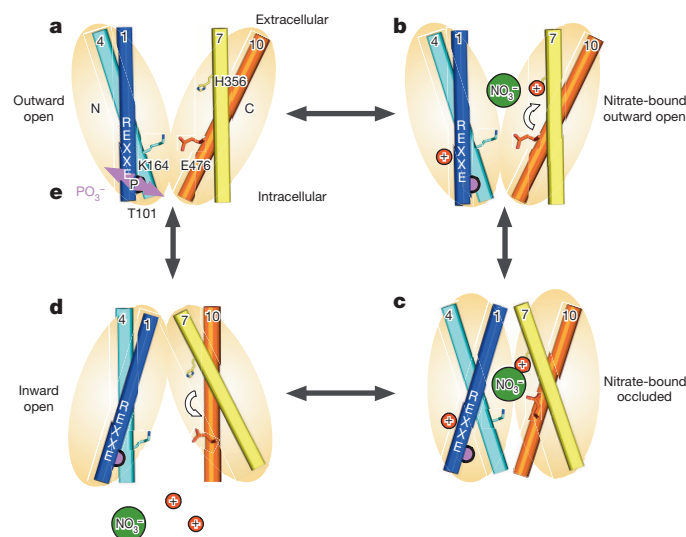


Figure 4 | Alternating access model for nitrate-proton symport. **a**, In the outward facing state the extracellular gate, constructed from TM1–TM2 (dark blue) and TM7–TM8 (yellow) is open and the intracellular gate, constructed from TM4–TM5 (cyan) packing against TM10–TM11 (orange) is closed and stabilized by a salt bridge between Lys 164 (TM4) and Glu 476 (TM10), similarly to the bacterial homologues. **b**, Following protonation of His 356, nitrate is able to bind triggering closure of the extracellular gate. Additional sites of protonation exist possibly within the ExxER motif (TM1). **c**, Upon transition to the occluded state the intracellular gate salt bridge will break as a result of Glu 476 moving to form a new salt bridge with His 356. **d**, This will result in the release of nitrate and protons into the interior of the cell. **e**, Phosphorylation of Thr 101 (purple sphere) by CIPK23 results in increased flexibility within the N-terminal bundle, increasing the overall transport rate.

the binding site. Perhaps the most surprising finding from our study, however, is that only minor changes to the previously characterized peptide transporter binding site are required to accommodate nitrate and that these are predominantly located in the C-terminal bundle. The ability of the PTR family to recognize multiple ligands may therefore stem from the separation of a fundamentally conserved proton coupling mechanism that resides largely in the N-terminal bundle and orchestrates gross conformational changes between TM4–TM5 and TM10–TM11, while leaving the C-terminal bundle to evolve to recognize different nitrogenous ligands.

METHODS SUMMARY

NRT1.1 was cloned into a modified, cleavable green fluorescent protein–histidine (GFP^{His}) fusion vector and expressed in *Saccharomyces cerevisiae*³⁵. The fusion protein was solubilized in 1% *n*-dodecyl- β -D-maltopyranoside (DDM) and purified to homogeneity. For MST binding measurements the GFP^{His} protein was left attached to the C terminus. For crystallization the GFP^{His} was removed through tobacco etch virus protease cleavage. Crystals were grown at pH 4.5 by the vapour diffusion method at 4 °C. Data were collected on beamlines IO2, IO3, IO4 and IO4-1 at the Diamond light source, UK. The crystals were derivatized with K₂Au(CN)₂ overnight and the phases calculated using single anomalous dispersion. For the nitrate co-complex, purification and crystallization was carried out in the presence of 10 mM sodium nitrate. The final apo and nitrate-bound structures were refined to 3.7 Å resolution. MST binding studies were carried out as described in methods following the standard protocols for the instrument. Binding isotherm data were fitted to a sigmoidal dose–response equation. Nitrate uptake was assayed in proteoliposomes containing wild-type, His356Ala, Thr101Asp variants of NRT1.1 or no protein controls. Nitrate uptake was measured as a function of coupled proton movement driven by an inwardly directed proton gradient through the quenching of the fluorescent pH indicator dye pyranine.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 October 2013; accepted 31 January 2014.

Published online 26 February 2014.

- Steiner, H.-Y., Naider, F. & Becker, J. M. The PTR family: a new group of peptide transporters. *Mol. Microbiol.* **16**, 825–834 (1995).
- Daniel, H., Spanier, B., Kottra, G. & Weitz, D. From bacteria to man: archaic proton-dependent peptide transporters at work. *Physiology (Bethesda)* **21**, 93–102 (2006).
- Crawford, N. M. Nitrate: nutrient and signal for plant growth. *Plant Cell* **7**, 859–868 (1995).
- Tsay, Y. F. & Hsu, P. K. The role of plasma membrane nitrogen transporters in nitrogen acquisition and utilization. *Plant Cell Monogr.* **19**, 223–236 (2011).
- Orsel, M. et al. Characterization of a two-component high-affinity nitrate uptake system in *Arabidopsis*: physiology and protein-protein interaction. *Plant Physiol.* **142**, 1304–1317 (2006).
- Dechorgnat, J. et al. From the soil to the seeds: the long journey of nitrate in plants. *J. Exp. Bot.* **62**, 1349–1359 (2011).
- Tsay, Y. F., Schroeder, J. I., Feldmann, K. A. & Crawford, N. M. The herbicide sensitivity gene CHL1 of *Arabidopsis* encodes a nitrate-inducible nitrate transporter. *Cell* **72**, 705–713 (1993).
- Léran, S. et al. A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends Plant Sci.* **19**, 5–9 (2013).
- Tsay, Y.-F., Chiu, C.-C., Tsai, C.-B., Ho, C.-H. & Hsu, P.-K. Nitrate transporters and peptide transporters. *FEBS Lett.* **581**, 2290–2300 (2007).
- Boursiac, Y. et al. ABA transport and transporters. *Trends Plant Sci.* **18**, 325–333 (2013).
- Nour-Eldin, H. H., Andersen, T. G., Burow, M. & Madsen, S. R. NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature* **488**, 531–534 (2012).
- Krouk, G. et al. Nitrate-regulated auxin transport by NRT1.1 defines a mechanism for nutrient sensing in plants. *Dev. Cell* **18**, 927–937 (2010).
- Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I. & Saier, M. H. The major facilitator superfamily (MFS) revisited. *FEBS J.* **279**, 2022–2035 (2012).
- Newstead, S. et al. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* **30**, 417–426 (2011).
- Fei, Y. J. et al. Expression cloning of a mammalian proton-coupled oligopeptide transporter. *Nature* **368**, 563–566 (1994).
- Nakajima, H. et al. Cloning and functional expression in *Escherichia coli* of the gene encoding the di- and tripeptide transport protein of *Lactobacillus helveticus*. *Appl. Environ. Microbiol.* **63**, 2213–2217 (1997).
- Chiang, C.-S., Stacey, G. & Tsay, Y.-F. Mechanisms and functional properties of two peptide transporters, AtPTR2 and fPTR2. *J. Biol. Chem.* **279**, 30150–30157 (2004).

18. Radestock, S. & Forrest, L. R. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J. Mol. Biol.* **407**, 698–715 (2011).
19. Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **38**, 151–159 (2013).
20. Madej, M. G. & Kaback, H. R. Evolutionary mix-and-match with MFS transporters. *Proc. Natl Acad. Sci. USA* **110**, E4831–E4838 (2013).
21. Solcan, N. *et al.* Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* **31**, 3411–3421 (2012).
22. Doki, S. *et al.* Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc. Natl Acad. Sci. USA* **110**, 11343–11348 (2013).
23. Liu, K. H., Huang, C. Y. & Tsay, Y. F. CHL1 is a dual-affinity nitrate transporter of *Arabidopsis* involved in multiple phases of nitrate uptake. *Plant Cell* **11**, 865–874 (1999).
24. Liu, K.-H. & Tsay, Y.-F. Switching between the two action modes of the dual-affinity nitrate transporter CHL1 by phosphorylation. *EMBO J.* **22**, 1005–1013 (2003).
25. Wang, Y.-Y., Hsu, P.-K. & Tsay, Y.-F. Uptake, allocation and signaling of nitrate. *Trends Plant Sci.* **17**, 458–467 (2012).
26. Baaske, P., Wienken, C., Willemsen, M. J. & Braun, D. Protein-binding assays in biological liquids using microscale thermophoresis. *J. Biomol. Tech.* **22**, S55 (2011).
27. Wienken, C. J., Baaske, P., Rothbauer, U. & Braun, D. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Commun.* **1**, 100 (2010).
28. Sun, L. *et al.* Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. *Nature* **490**, 361–366 (2012).
29. Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467**, 734–738 (2010).
30. Guettou, F. *et al.* Structural insights into substrate recognition in proton-dependent oligopeptide transporters. *EMBO Rep.* **14**, 804–810 (2013).
31. Yan, H. *et al.* Structure and mechanism of a nitrate transporter. *Cell Rep.* **3**, 716–723 (2013).
32. Chen, X.-Z., Zhu, T., Smith, D. E. & Hediger, M. A. Stoichiometry and kinetics of the high-affinity H⁺-coupled peptide transporter PepT2. *J. Biol. Chem.* **274**, 2773–2779 (1999).
33. Chandrasekaran, A., Ojeda, A. M., Kolmakova, N. G. & Parsons, S. M. Mutational and bioinformatics analysis of proline- and glycine-rich motifs in vesicular acetylcholine transporter. *J. Neurochem.* **98**, 1551–1559 (2006).
34. Ugolev, Y., Segal, T., Yaffe, D., Gros, Y. & Schuldiner, S. Identification of conformationally sensitive residues essential for inhibition of vesicular monoamine transport by the noncompetitive inhibitor-tetrabenazine. *J. Biol. Chem.* **288**, 32160–32171 (2013).
35. Newstead, S. *et al.* High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **104**, 13936–13941 (2007).

Acknowledgements We thank R. Flaig for help with additional access to beamline I04 and we thank Diamond Light Source for access to beam lines I02, I03, I04 and I04-1 (MX7345) that contributed to the results presented here. This research was funded through the Medical Research Council (MRC) Career Development Award grant G0900399 and Royal Society grants (RG110211 and IE111401) to S.N.

Author Contributions J.L.P. and S.N. designed, performed and analysed all experiments and wrote the manuscript.

Author Information The atomic coordinates and structure factors of the apo and nitrate bound NRT1.1 are deposited in the Protein Data Bank with accession codes 4cl4 and 4cl5, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L.P. (joanne.parker@bioch.ox.ac.uk) or S.N. (simon.newstead@bioch.ox.ac.uk).

Crystal structure of the plant dual-affinity nitrate transporter NRT1.1

Ji Sun¹, John R. Bankston², Jian Payandeh^{1†}, Thomas R. Hinds¹, William N. Zagotta² & Ning Zheng^{1,3}

Nitrate is a primary nutrient for plant growth, but its levels in soil can fluctuate by several orders of magnitude. Previous studies have identified *Arabidopsis* NRT1.1 as a dual-affinity nitrate transporter that can take up nitrate over a wide range of concentrations. The mode of action of NRT1.1 is controlled by phosphorylation of a key residue, Thr 101; however, how this post-translational modification switches the transporter between two affinity states remains unclear. Here we report the crystal structure of unphosphorylated NRT1.1, which reveals an unexpected homodimer in the inward-facing conformation. In this low-affinity state, the Thr 101 phosphorylation site is embedded in a pocket immediately adjacent to the dimer interface, linking the phosphorylation status of the transporter to its oligomeric state. Using a cell-based fluorescence resonance energy transfer assay, we show that functional NRT1.1 dimerizes in the cell membrane and that the phosphomimetic mutation of Thr 101 converts the protein into a monophasic high-affinity transporter by structurally decoupling the dimer. Together with analyses of the substrate transport tunnel, our results establish a phosphorylation-controlled dimerization switch that allows NRT1.1 to uptake nitrate with two distinct affinity modes.

Active nitrate (NO_3^-) uptake by roots represents the critical first step of nitrogen acquisition in plants, which render the essential element to animals in organic forms. The abundance of nitrate in soil is affected by many environmental factors. As a result, soil concentrations of nitrate can undergo rapid changes, varying from low micromolar to high millimolar concentrations. In adaptation to fluctuating nitrate levels, plants have evolved two complementary nitrate transporter systems with distinct kinetic properties^{1–3}. The high-affinity transport system (HATS), which consists of members of the NRT2/NNP family, drives nitrate uptake with Michaelis constant (K_m) values in the micromolar range, whereas the low-affinity transport system (LATS), comprising the NRT1/PTR family, transports nitrate at millimolar concentrations. In response to changes in nitrogen availability and demands, the activity of select components of these two systems can be further fine-tuned by transcriptional regulation and post-translational modifications. Transporters of both families belong to the major facilitator superfamily (MFS) of secondary active transporters and are dependent on protons for nitrate transport^{4,5}.

The *Arabidopsis* NRT1.1 protein (also called CHL1 or NPF6.3) is the first nitrate transporter identified in higher plants and belongs to the NRT1 family^{6,7}. Distinct from most of the NRT1 and NRT2 family members, NRT1.1 functions as a dual-affinity transporter and contributes to both HATS and LATS^{8–10}. In comparison to the wild-type plant, *nrt1.1* mutants show marked nitrate uptake defects in both high- and low-affinity ranges. In the heterologous *Xenopus* oocyte expression system, the transporter activity of NRT1.1 exhibits a characteristic biphasic kinetics with two different K_m values of $\sim 50 \mu\text{M}$ and $\sim 4 \text{mM}$. Notably, recent studies have shown that phosphorylation of a single residue, Thr 101, is responsible for switching NRT1.1 from the low-affinity to the high-affinity mode¹¹. Mutations of Thr 101 preventing or mimicking phosphorylation can effectively convert the dual-affinity transporter into a monophasic low-affinity or high-affinity transporter, respectively. Nitrate is not only a nutrient but also a signalling molecule, which modulates many aspects of plant physiology and optimal nitrate

acquisition^{12–15}. Independent of its transporter function, NRT1.1 also acts as a nitrate sensor, regulating differential expression of primary nitrate response genes at different nitrate levels^{16–18}. Remarkably, the sensor function of NRT1.1 also shows a biphasic pattern and is affected by the phosphorylation status of Thr 101.

The MFS permeases transport a wide spectrum of substrates and represent one of the largest families of secondary carriers in all species of life. Despite the growing number of available crystal structures of MFS transporters^{19–29}, little is known about the structural mechanisms regulating their activities. Here we report the crystal structure of the full-length *Arabidopsis thaliana* NRT1.1, which reveals an unexpected phosphorylation-controlled dimerization switch that enables the transporter to operate with a dual-affinity mode.

Overall structure of NRT1.1

The *Arabidopsis thaliana* NRT1.1 gene encodes a 590-amino-acid protein, which is highly conserved among plant NRT1.1 orthologues, but not *Arabidopsis* NRT1 family members (Extended Data Figs 1 and 2). The recombinant NRT1.1 protein was overexpressed, solubilized and purified from insect cells with dodecyl maltoside (DDM) and crystallized in the presence of 10 mM NaNO_3 . With combined phases from Rosetta-improved molecular replacement³⁰ and single wavelength anomalous diffraction, we determined and refined the NRT1.1 structure at 3.25 Å resolution (Extended Data Fig. 3 and Extended Data Table 1).

NRT1.1 crystallized with two molecules in the asymmetric unit. The two copies of NRT1.1 can be superimposed with a root mean square deviation of 0.9 Å over 504 C α atoms, indicating a common overall structure (Fig. 1a). As predicted, the transporter adopts a typical MFS fold, which is characterized by 12 transmembrane helices (TMHs) with a pseudo two-fold axis relating the amino-terminal (TMH1–6) and carboxy-terminal (TMH7–12) domains (Fig. 1b). NRT1.1 is captured in an inward-facing conformation as previously observed in the LacY (ref. 19), GlpT (ref. 20) and PepT_{St} (ref. 25) structures (Fig. 1c and Extended Data Fig. 4).

¹Department of Pharmacology, Box 357280, University of Washington, Seattle, Washington 98195, USA. ²Department of Physiology and Biophysics, Box 357290, University of Washington, Seattle, Washington 98195, USA. ³Howard Hughes Medical Institute, Box 357280, University of Washington, Seattle, Washington 98195, USA. [†]Present address: Department of Structural Biology, Genentech Inc., South San Francisco, California 94080, USA.

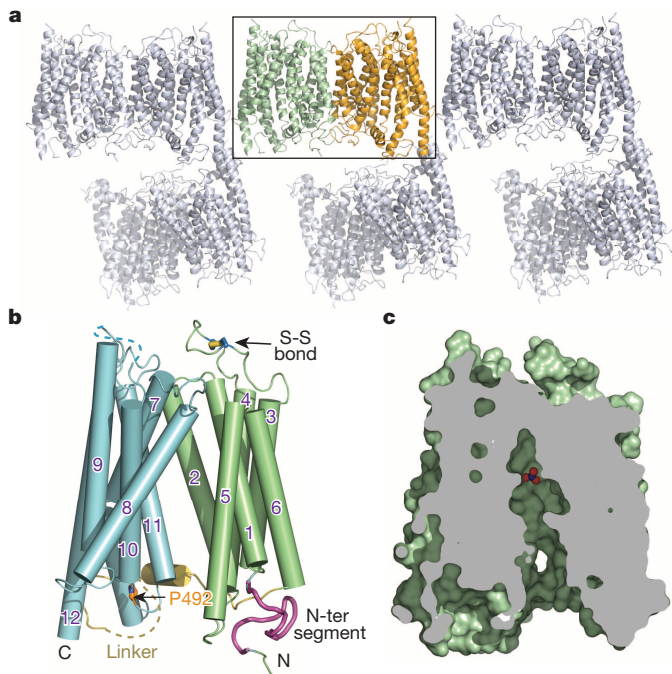


Figure 1 | Crystal packing and overall structure of NRT1.1. **a**, Crystal packing of NRT1.1 in space group C2221 with two molecules in each asymmetric unit. **b**, Overall structure of NRT1.1. The N-terminal and C-terminal domains, the N-terminal conserved segment, the inter-domain linker and Pro 492 are coloured in pale green, cyan, magenta, yellow and orange, respectively. A functional important extracellular disulphide bond is indicated. **c**, Cutaway view showing that NRT1.1 is captured in an inward conformation with nitrate displayed as spheres.

By comparing the structures of NRT1.1 and a bacterial peptide transporter from *Streptococcus thermophilus*, PepT_{ST}, we confirm that eukaryotic and prokaryotic members of the NRT1/PTR family of MFS transporters share a similar overall architecture (Extended Data Fig. 4). The plant nitrate transporter, nonetheless, has three unique and conserved structural elements, including a well-structured N-terminal cytoplasmic segment, a disulphide bond-stabilized extracellular loop, and a partially ordered central linker sequence (Fig. 1b). Consistent with the sequence divergence between the NRT1/PTR and NRT2/NNP nitrate transporter families, the structures of NRT1.1 and two bacterial NRT2/NNP family members, NarK (ref. 24) and NarU (ref. 26), share few common features except for the MFS fold (Extended Data Fig. 4).

The residue mutated (P492L) in the *chl1-9* mutant, which lost the transporter but not the sensor function of NRT1.1 (ref. 17), is located at the short TMH10–TMH11 loop (Fig. 1b). Its mutation probably affects the structural coordination of the two helices.

NRT1.1 dimer in the crystal

So far, crystal structures of more than ten MFS transporters have been determined in the monomeric form^{19–29}. The DDM-solubilized NRT1.1 protein was also isolated in a monomeric state as determined by size-exclusion chromatography-coupled multi-angle light scattering measurements³¹ (Extended Data Fig. 5a). A closer examination of the two NRT1.1 molecules in the asymmetric unit, however, reveals a possible biological dimer.

In the crystal, the two adjacent non-crystallographically related NRT1.1 molecules are juxtaposed in a side-to-side fashion with their N-terminal halves facing and interacting with each other (Figs 1a and 2a, b). The intermolecular packing is predominantly mediated by TMH3 and TMH6, which are located at a peripheral edge of the canonical MFS fold. Although crystal contacts may not always reflect biological interactions, two prominent features of the crystallographic dimer arrangement

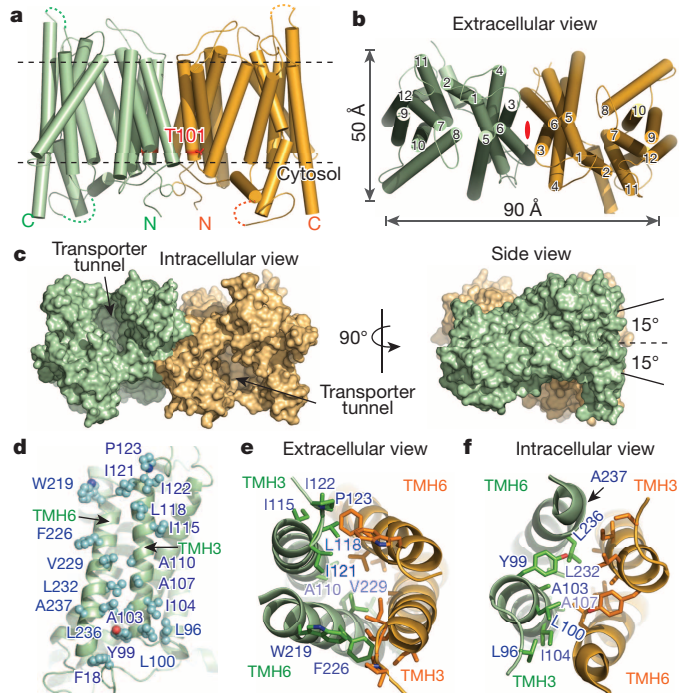


Figure 2 | NRT1.1 dimer interface. **a**, Cylinder representation of the NRT1.1 dimer with Thr 101 shown as red sticks. **b**, Extracellular view of the NRT1.1 dimer with a central two-fold axis indicated in red. **c**, Two orthogonal views of the NRT1.1 dimer in surface representation. The dashed line represents the central two-fold axis. **d**, NRT1.1 dimer interface. The side chains of all interface residues are shown as spheres. **e**, **f**, Extracellular and intracellular views of the NRT1.1 dimer interface with TMH3 and TMH6 shown in ribbon representation and the side chains of interacting residues shown as sticks. The interface residues of one chain are labelled.

support its physiological relevance. First, the overall topology of the putative NRT1.1 dimer is perfectly compatible with its transporter function at the membrane (Fig. 2a, b). Second, the interface between the two NRT1.1 molecules is extensive and complementary with a total surface area of $\sim 2,160 \text{ \AA}^2$ (Fig. 2d–f and Extended Data Fig. 6). Overall, the two inward-facing NRT1.1 molecules give rise to a putative ‘in-phase’ dimer assembly, which is about 90 Å wide and 50 Å thick (Fig. 2b). When viewed from the side, the two substrate-transporting tunnels are not in parallel with the central two-fold axis but slant at an $\sim 15^\circ$ angle in opposite directions (Fig. 2c).

Functional dimerization of NRT1.1

To dissect the biological relevance of the NRT1.1 dimer observed in the crystal, we first used a crosslinking experiment to assess the potential of detergent-solubilized NRT1.1 to dimerize in solution. Despite its low efficiency, an amine reactive crosslinker was able to crosslink NRT1.1 in a concentration-dependent manner (Fig. 3a). The cross-linked products migrated on SDS–polyacrylamide gel electrophoresis (PAGE) with a size corresponding to a NRT1.1 dimer, indicating that DDM-solubilized NRT1.1 is capable of forming a transient dimer in a membrane-free environment.

Because solubilization by DDM might interfere with NRT1.1 dimer formation, we next performed fluorescence resonance energy transfer (FRET) spectroscopy experiments with the nitrate transporter expressed in the membrane of *Xenopus* oocytes³², which allowed us to examine the oligomeric state of NRT1.1 in the same lipid environment where its dual-affinity transporter activity has been measured. We separately fused the N terminus of NRT1.1 with either the mCerulean variant of cyan fluorescent protein (mCFP) or the mCitrine variant of yellow fluorescent protein (mYFP), which constitute a FRET pair with an R_0 of $\sim 50 \text{ \AA}$ for 50% energy transfer efficiency³³ (Fig. 3b). In the structure

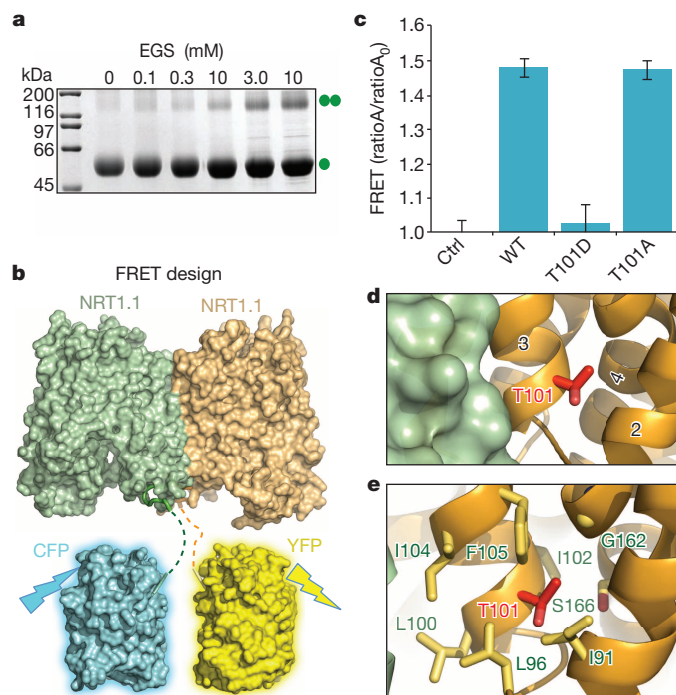


Figure 3 | NRT1.1 dimerization controlled by Thr 101 phosphorylation. **a**, Crosslinking of NRT1.1 with increasing concentrations of ethylene glycol bis-succinimidylsuccinate (EGS). **b**, The design of FRET assay. Dashed lines indicate the 11-residue-long linkers between the fluorescence proteins and the structurally resolved NRT1.1 N terminus. **c**, FRET measurements of wild-type (WT) and mutant NRT1.1. The mCFP–HCN–mYFP–NRT1.1 pair was used as negative control. Consistent with the loss of FRET signal, the T101D mutant failed to be crosslinked in solution (Extended Data Fig. 5b). **d**, A close-up view of Thr 101 at the NRT1.1 dimer interface. **e**, Thr 101-interacting residues with their side chains shown as sticks.

of the putative NRT1.1 dimer, the N-terminal ends of the two NRT1.1 molecules are about 40 Å away from each other (Extended Data Fig. 7a). Therefore, FRET is expected to occur if NRT1.1 dimerizes in the membrane in the same fashion as seen in the crystal structure. As shown in Fig. 3c, a strong FRET signal measured by a spectrum-based approach was detected between the co-expressed mCFP–NRT1.1 and mYFP–NRT1.1 fusion proteins, but not in the negative control (Extended Data Fig. 7b). This result strongly suggests that the plant dual-affinity nitrate transporter can form a homodimer not only in the crystal but also in the cellular membrane.

Thr 101 phosphorylation as a dimerization switch

The phosphorylation site residue Thr 101 is strictly conserved among plant NRT1.1 orthologues and represents one of the hallmarks of the dual-affinity nitrate transporter¹¹ (Extended Data Fig. 1). In the NRT1.1 structure, Thr 101 is located at the N-terminal end of TMH3 and is entirely buried in a hydrophobic pocket formed among TMH2, TMH3 and THM4 (Fig. 3d, e). Notably, this pocket is directly adjacent to the dimer interface with one of its walls demarcated by three hydrophobic interface residues: Leu 96, Leu 100 and Ile 104 (Figs 2d and 3e). Although the side chain of Thr 101 is unmodified in the crystal, its phosphorylation is expected to induce major electrostatic and conformational changes in its vicinity and have a direct impact on the dimer interface. This structural feature prompted us to postulate that the formation of the NRT1.1 dimer might be determined by the phosphorylation status of Thr 101 and that the two distinct affinity states of the transporter might be enabled by the difference in its oligomerization state.

To test this hypothesis, we compared the wild-type NRT1.1 with the phosphomimetic mutant, T101D, and the phosphorylation-defective

mutant, T101A, in the oocyte-based FRET experiments (Fig. 3c and Extended Data Fig. 7b). Consistent with the prediction, the phosphomimetic mutant T101D, which has been previously shown to bear a monophasic high-affinity nitrate transporter activity¹¹, completely lost the FRET signal, indicating a spatial separation of the two NRT1.1 N-terminal domains, if not a complete disruption of the NRT1.1 dimer. By contrast, the phosphorylation-defective mutant T101A, which is known to transport nitrate in the low-affinity state¹¹, generated a robust FRET signal similar to the wild-type protein. The matching levels of FRET between wild-type NRT1.1 and the T101A mutant indicates that the wild-type protein is mostly in the unmodified form under the nitrate-free condition of the FRET measurement. This is in agreement with previous studies showing that phosphorylated NRT1.1 only started to accumulate when plants were exposed to nitrate. Together, these results not only confirm the functional relevance of the NRT1.1 dimer, but also indicate a dimerization-based switching mechanism for the dual-affinity nitrate transporter—unmodified NRT1.1 forms a structurally coupled homodimer and functions as a low-affinity transporter, whereas phosphorylated NRT1.1 undergoes dimer decoupling and adopts a high-affinity state.

Substrate binding site and proton coupling

In both protomers of the refined NRT1.1 dimer structure, an island of strong electron density is present in the middle of the transport tunnel between the N-terminal and C-terminal domains (Fig. 4a and Extended Data Fig. 8). The overall location of the density coincides with the substrate-binding sites of other MFS structures, indicating that it belongs to the substrate molecule, nitrate. Indeed, when nitrate was

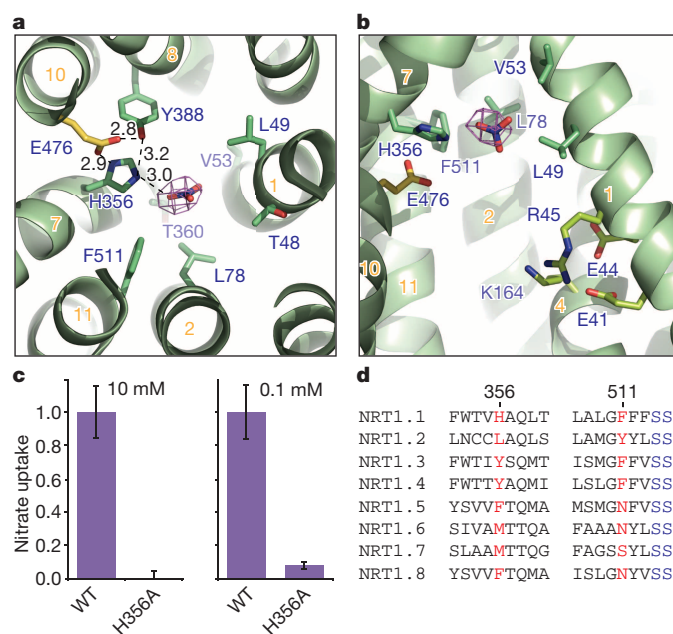


Figure 4 | Substrate binding and energy coupling in NRT1.1. **a**, Intracellular view of the nitrate-binding pocket. Nitrate is shown as a stick representation together with electron density contoured at 4σ from a $F_o - F_c$ map calculated before the nitrate was modelled in. THMs are numbered (orange). E476 is a His 356-interacting residue, the mutation of which abolishes the transporter function of NRT1.1 (Extended Data Fig. 3b). **b**, Side view of the putative nitrate-binding site and the transporter tunnel with the clustered EXXER motif and K164. **c**, Nitrate uptake activities of the H356A mutant relative to wild-type NRT1.1 in the presence of 10 mM or 0.1 mM nitrate. y axis is the percentage of nitrate uptake compared to wild type. All results are the mean \pm s.d. of one experiment in quintuplicates or sextuplicates. **d**, Sequence alignment of eight NRT1 family members from *Arabidopsis thaliana* in regions surrounding H356 and F511 of NRT1.1.

omitted from the cryo-protection buffer, this electron density completely disappeared from both NRT1.1 molecules (Extended Data Fig. 8). The position of the nitrate density is slightly shifted between the two NRT1.1 molecules, which might reflect the mode of substrate release (Extended Data Fig. 8).

Distinct from the nitrate/nitrite binding sites of NarK and NarU, which coordinate the substrate(s) with two opposing conserved Arg residues^{24,26}, the nitrate-binding pocket in NRT1.1 is predominantly formed by hydrophobic residues, including Leu 49, Val 53, Leu 78 and Phe 511 (Fig. 4a, b). His 356 on TMH7 is the only polar residue that is in close contact with nitrate, the precise binding mode of which cannot be resolved due to the resolution limit of the structure. On the basis of its close proximity to the nitrate density and the crystallization condition (pH = 4.5), His 356 probably stabilizes the substrate in the pocket through a charge–charge interaction. Its side-chain conformation, meanwhile, is supported by two nearby residues: Tyr 388 and Glu 476. Although Tyr 388 and two other polar residues, Thr 360 and Thr 48, are also around the substrate, their hydroxyl groups do not seem to be at the optimal hydrogen bond distance (Fig. 4a).

To validate the substrate-binding site, we mutated His 356 and compared the nitrate uptake activities of the wild-type and mutant transporters¹¹. In support of a critical role of His 356 in substrate transport, its mutation to alanine completely abolished the transport activity of NRT1.1 at both high and low nitrate concentration (Fig. 4c). Notably, His 356 is not conserved among plant NRT1.1 orthologues and *Arabidopsis* NRT1 family members, which harbour either a tyrosine or a hydrophobic amino acid (Leu, Met or Phe) at the equivalent position (Fig. 4d). This key residue, nevertheless, has closely co-evolved with the adjacent residue, Phe 511. Among all NRT1.1 orthologues and paralogues, a combination of a polar and a hydrophobic side chain has been generally maintained between the two residues, indicating that one of them is responsible for specific nitrate binding (Fig. 4d and Extended Data Figs 1 and 2). NRT1.1 is unique among all *Arabidopsis* NRT1 family members by featuring a histidine at the nitrate-binding pocket. This charged residue provides a plausible explanation for the high-affinity nitrate uptake activity acquired by NRT1.1, which is otherwise a member of LATS. Furthermore, the replacement of the histidine residue by tyrosine in some of the plant NRT1.1 orthologues raises a question about their dual affinity transporter function (Extended Data Fig. 1).

In the PepT_{St} structure, a conserved motif, EXXERFXYY, on TMH1 has been identified to have an important role in proton coupling. Part of this motif, EXXER, is also found in all plant NRT1.1 orthologues (Extended Data Fig. 1). Together with the conserved residue Lys 164, this motif presents a cluster of interacting residues under the nitrate-binding pocket and facing towards the transport tunnel (Fig. 4b). Consistent with a key function in the symport cycle, alanine mutation of each of the four residues abrogated the transporter activity of NRT1.1 in the oocyte-based nitrate uptake assay (Extended Data Fig. 3b). Surprisingly, these four residues have been simultaneously evolved into non-charged residues in two *Arabidopsis* NRT1 family members: AtNRT1.5 and AtNRT1.8 (Extended Data Fig. 2). Their documented pH-dependent nitrate transporter activities necessitate an alternative proton-coupling mechanism.

Cytoplasmic structural elements

NRT1.1 has an ~30-amino-acid-long N-terminal cytoplasmic segment, which is highly conserved among its plant orthologues (Extended Data Figs 1 and 2). In the crystal, this sequence adopts a well-ordered loop structure and forms a pronounced cleft between the NRT1.1 dimer (Extended Data Fig. 9). With several strictly conserved residues exposed to the solvent, this cleft presents a putative two-fold symmetric protein–protein interaction site with a potential role in recruiting kinases and phosphatases. Although the central linker sequence is mostly disordered in the crystal, its N-terminal region forms a stable amphipathic α -helix (Figs 1b and 2a), providing yet another potential protein-docking site.

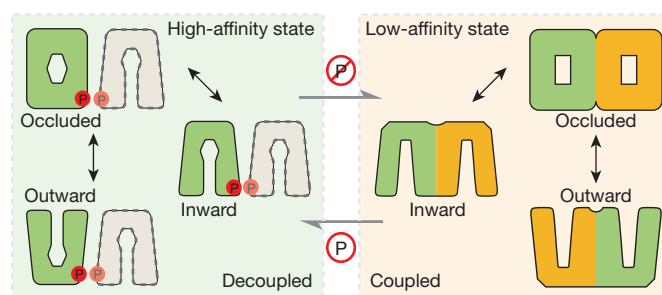


Figure 5 | A dimerization switch model. The non-phosphorylated and structurally coupled NRT1.1 dimer functions as an ‘in-phase’ homodimeric low-affinity nitrate transporter (right). Once phosphorylated, the NRT1.1 dimer is decoupled, and each molecule functions as an independent high-affinity nitrate transporter (left). Different shapes of the putative substrate-binding site at the central transport tunnel reflect its differential nitrate-binding properties.

Discussion

The crystal structure of NRT1.1 reveals a biologically relevant dimer, the dynamic coupling and decoupling of which is controlled by the phosphorylation of a single residue near the dimer interface. Because the same post-translational modification switches the mode of action of the dual-affinity transporter, we propose that dimer assembly and disassembly enables NRT1.1 to toggle between the low-affinity and high-affinity states with an overlapping, if not the same, substrate-binding site. In this model (Fig. 5), structural engagement of two protomers at the interface allosterically regulates the affinity of substrate binding at the central transport tunnel. Whether dimer decoupling itself is sufficient to shift NRT1.1 into the high-affinity mode awaits future analysis.

Previous structural studies of several MFS members have established a ‘rocker-switch’ mechanism of substrate transport, in which the transporters cycle through outward-facing, occluded, and inward-facing conformations^{21,25}. The structure of the dimeric unmodified NRT1.1 reveals a buried Thr 101 phosphorylation site in the inward-facing conformation. Thr 101 phosphorylation, therefore, probably occurs when the dimeric transporter adopts either the outward-facing or occluded conformation. It is equally possible that the unmodified dimeric transporter is in equilibrium with the monomeric form, which is susceptible to phosphorylation.

Despite the current structure, questions remain as to how NRT1.1 senses nitrate and transduces the signal. Our structure reveals only one nitrate-binding site within the substrate transport tunnel. However, there might be an additional nitrate-binding site responsible for signalling, which is excluded from the inward-facing conformation. Previous studies have suggested that dephosphorylation of Thr 101 is required for the low-affinity sensor function of NRT1.1 (ref. 17). It is possible that dimerization mediates this signalling function of the transporter in the same manner as found in many common cell-surface receptors. The high-affinity sensor function of NRT1.1, on the other hand, might involve an entirely different mechanism.

Transporter oligomerization and phosphorylation have been implicated in the proper functions of a number of MFS members, such as LacS (ref. 34), GLUT1 (ref. 35), TetL (ref. 36) and hRFC (ref. 37). The crystal structure of the NRT1.1 dimer not only establishes a structural framework for understanding its dual-affinity nitrate transporter and receptor activities, but also reveals how protein oligomerization and post-translational modification can synergistically expand the functional capacity of an MFS transporter.

METHODS SUMMARY

Detailed descriptions of the following experimental procedures can be found in the Methods: protein expression and purification; protein crystallization, data collection and structure determination; nitrate transporter assay in oocytes; FRET assay and data analysis; and details for crosslinking experiment.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 September 2013; accepted 23 January 2014.

Published online 26 February 2014.

- Wang, Y. Y., Hsu, P. K. & Tsay, Y. F. Uptake, allocation and signaling of nitrate. *Trends Plant Sci.* **17**, 458–467 (2012).
- Tsay, Y. F., Chiu, C. C., Tsai, C. B., Ho, C. H. & Hsu, P. K. Nitrate transporters and peptide transporters. *FEBS Lett.* **581**, 2290–2300 (2007).
- Nacry, P. B. & Gojon, E. A. Nitrogen acquisition by roots: physiological and developmental mechanisms ensuring plant adaptation to a fluctuating resource. *Plant Soil* **370**, 1–29 (2013).
- Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
- Law, C. J., Maloney, P. C. & Wang, D. N. Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **62**, 289–305 (2008).
- Leran, S. *et al.* A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends Plant Sci.* **19**, 5–9 (2013).
- Tsay, Y. F., Schroeder, J. I., Feldmann, K. A. & Crawford, N. M. The herbicide sensitivity gene CHL1 of *Arabidopsis* encodes a nitrate-inducible nitrate transporter. *Cell* **72**, 705–713 (1993).
- Huang, N. C., Chiang, C. S., Crawford, N. M. & Tsay, Y. F. CHL1 encodes a component of the low-affinity nitrate uptake system in *Arabidopsis* and shows cell type-specific expression in roots. *Plant Cell* **8**, 2183–2191 (1996).
- Wang, R., Liu, D. & Crawford, N. M. The *Arabidopsis* CHL1 protein plays a major role in high-affinity nitrate uptake. *Proc. Natl Acad. Sci. USA* **95**, 15134–15139 (1998).
- Liu, K. H., Huang, C. Y. & Tsay, Y. F. CHL1 is a dual-affinity nitrate transporter of *Arabidopsis* involved in multiple phases of nitrate uptake. *Plant Cell* **11**, 865–874 (1999).
- Liu, K. H. & Tsay, Y. F. Switching between the two action modes of the dual-affinity nitrate transporter CHL1 by phosphorylation. *EMBO J.* **22**, 1005–1013 (2003).
- Guo, F. Q., Young, J. & Crawford, N. M. The nitrate transporter AtNRT1.1 (CHL1) functions in stomatal opening and contributes to drought susceptibility in *Arabidopsis*. *Plant Cell* **15**, 107–117 (2003).
- Wang, R., Okamoto, M., Xing, X. & Crawford, N. M. Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* **132**, 556–567 (2003).
- Krouk, G. *et al.* Nitrate-regulated auxin transport by NRT1.1 defines a mechanism for nutrient sensing in plants. *Dev. Cell* **18**, 927–937 (2010).
- Walch-Liu, P. *et al.* Nitrogen regulation of root branching. *Ann. Bot.* **97**, 875–881 (2006).
- Munos, S. *et al.* Transcript profiling in the chl1–5 mutant of *Arabidopsis* reveals a role of the nitrate transporter NRT1.1 in the regulation of another nitrate transporter, NRT2.1. *Plant Cell* **16**, 2433–2447 (2004).
- Ho, C. H., Lin, S. H., Hu, H. C. & Tsay, Y. F. CHL1 functions as a nitrate sensor in plants. *Cell* **138**, 1184–1194 (2009).
- Bouguyon, E., Gojon, A. & Nacry, P. Nitrate sensing and signaling in plants. *Semin. Cell Dev. Biol.* **23**, 648–654 (2012).
- Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
- Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301**, 616–620 (2003).
- Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467**, 734–738 (2010).
- Newstead, S. *et al.* Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* **30**, 417–426 (2011).
- Pedersen, B. P. *et al.* Crystal structure of a eukaryotic phosphate transporter. *Nature* **496**, 533–536 (2013).
- Zheng, H., Wisedchaisri, G. & Gonen, T. Crystal structure of a nitrate/nitrite exchanger. *Nature* **497**, 647–651 (2013).
- Solcan, N. *et al.* Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* **31**, 3411–3421 (2012).
- Yan, H. *et al.* Structure and mechanism of a nitrate transporter. *Cell Rep.* **3**, 716–723 (2013).
- Sun, L. *et al.* Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. *Nature* **490**, 361–366 (2012).
- Yin, Y., He, X., Szweczyk, P., Nguyen, T. & Chang, G. Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science* **312**, 741–744 (2006).
- Doki, S. *et al.* Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc. Natl Acad. Sci. USA* **110**, 11343–11348 (2013).
- DiMaio, F. *et al.* Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* **473**, 540–543 (2011).
- Slotboom, D. J., Duurkens, R. H., Olieman, K. & Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46**, 73–82 (2008).
- Zheng, J., Trudeau, M. C. & Zagotta, W. N. Rod cyclic nucleotide-gated channels have a stoichiometry of three CNGA1 subunits and one CNGB1 subunit. *Neuron* **36**, 891–896 (2002).
- Taraska, J. W. & Zagotta, W. N. Fluorescence applications in molecular neurobiology. *Neuron* **66**, 170–189 (2010).
- Veenhoff, L. M., Heuberger, E. H. & Poolman, B. The lactose transport protein is a cooperative dimer with two sugar translocation pathways. *EMBO J.* **20**, 3056–3062 (2001).
- Pessino, A. *et al.* Evidence that functional erythrocyte-type glucose transporters are oligomers. *J. Biol. Chem.* **266**, 20213–20217 (1991).
- Safferling, M. *et al.* TetL tetracycline efflux protein from *Bacillus subtilis* is a dimer in the membrane and in detergent solution. *Biochemistry* **42**, 13969–13976 (2003).
- Hou, Z., Cherian, C., Drews, J., Wu, J. & Matherly, L. H. Identification of the minimal functional unit of the homo-oligomeric human reduced folate carrier. *J. Biol. Chem.* **285**, 4732–4740 (2010).

Acknowledgements We thank the beamline staff of the Advanced Light Source at the University of California at Berkeley. We also thank members of the Zheng laboratory, Zagotta laboratory, Xu laboratory and H. Zheng for discussion and help. This work is supported by the Howard Hughes Medical Institute (N. Z.), National Institutes of Health (R01EY10329 to W.N.Z., NS074545 to J.R.B.) and the National Science Foundation (N.Z.).

Author Contributions J.S. and N.Z. conceived and J.S. conducted the protein purification and crystallization experiments. J.P. provided experimental suggestions. J.S. and N.Z. determined and analysed the structures. J.S., T.R.H. and N.Z. conceived and J.S. and T.R.H. conducted SEC-LS-RI-UV experiments. J.S., J.R.B., W.N.Z. and N.Z. conceived and J.S. and J.R.B. conducted FRET experiments. J.S. and N.Z. conceived and J.S. conducted mutational and transporter assays. J.S. and N.Z. wrote the manuscript with inputs from all authors.

Author Information Structural coordinates and structural factors are deposited in the Protein Data Bank under accession number 4OH3. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.Z. (nzheng@u.washington.edu).

Change in the chemical composition of infalling gas forming a disk around a protostar

Nami Sakai¹, Takeshi Sakai², Tomoya Hirota³, Yoshimasa Watanabe¹, Cecilia Ceccarelli⁴, Claudine Kahane⁴, Sandrine Bottinelli^{5,6}, Emmanuel Caux^{5,6}, Karine Demyk^{5,6}, Charlotte Vastel^{5,6}, Audrey Coutens^{7,8}, Vianney Taquet⁹, Nagayoshi Ohashi^{10,11}, Shigehisa Takakuwa¹², Hsi-Wei Yen^{10,12}, Yuri Aikawa¹³ & Satoshi Yamamoto¹

IRAS 04368+2557 is a solar-type (low-mass) protostar embedded in a protostellar core (L1527) in the Taurus molecular cloud^{1,2}, which is only 140 parsecs away from Earth, making it the closest large star-forming region. The protostellar envelope has a flattened shape with a diameter of a thousand astronomical units (1 AU is the distance from Earth to the Sun), and is infalling and rotating^{3–5}. It also has a protostellar disk with a radius of 90 AU (ref. 6), from which a planetary system is expected to form^{7,8}. The interstellar gas, mainly consisting of hydrogen molecules, undergoes a change in density of about three orders of magnitude as it collapses from the envelope into the disk, while being heated from 10 kelvin to over 100 kelvin in the mid-plane, but it has hitherto not been possible to explore changes in chemical composition associated with this collapse. Here we report that the unsaturated hydrocarbon molecule cyclic-C₃H₂ resides in the infalling rotating envelope, whereas sulphur monoxide (SO) is enhanced in the transition zone at the radius of the centrifugal barrier (100 ± 20 AU), which is the radius at which the kinetic energy of the infalling gas is converted to rotational energy. Such a drastic change in chemistry at the centrifugal barrier was not anticipated, but is probably caused by the discontinuous infalling motion at the centrifugal barrier and local heating processes there.

We have conducted high-spatial-resolution observations of the millimetre and submillimetre wave lines of cyclic-C₃H₂ and SO in the 500 AU region around IRAS 04368+2557 (Methods). Figure 1a shows the observed integrated-intensity map of the 5_{23–432} rotational line of cyclic-C₃H₂ (where $J_{KK'}$ represents the rotational energy level of an asymmetric top molecule; see the footnote of Extended Data Table 1). The intensity distribution shows a double-peaked structure, where the southern and northern peaks are separated from the protostar position (the continuum peak position: Extended Data Fig. 1) by 1'' (140 AU). Figure 1b shows the position–velocity (PV) diagram along the north–south line centred at the protostar's position. A rotation signature is clearly seen outward of 100 AU from the protostar, where the brightest velocity component is redshifted and blueshifted for the northern and southern parts, respectively. This situation is schematically illustrated in Fig. 1c. Such a rotation signature abruptly disappears inward of 100 AU from the protostar. In addition to this signature, weak blueshifted and redshifted components can also be recognized in the northern and southern parts, respectively (Fig. 1b). The existence of such counter-velocity components indicates that the gas motion is not simple rotation but accompanies infalling motion, as shown in Fig. 2a. Most importantly, only the infalling (counter-velocity) components are seen towards the protostar position. This means that cyclic-C₃H₂ is almost absent inward of 100 AU. The other cyclic-C₃H₂ lines (5_{51–440}, 9_{18–827}/9_{28–817} and

10_{0,10–919}/10_{1,10–909}) also show similar distributions (Extended Data Fig. 2).

To account for this kinematic feature, we developed a toy model describing an infalling rotating envelope, where the gas motion is approximated by the motion of a particle for simplicity (Fig. 2a) (Methods and Extended Data Fig. 3). According to this model, the maximum infalling velocity towards the protostar position is half of the maximum rotation velocity at the centrifugal barrier of the infalling gas (that is, half the centrifugal radius). In fact, the observed maximum infalling velocity of cyclic-C₃H₂ towards the protostar position is just half of the maximum rotation velocity (Fig. 1b). Therefore, the radius that gives the maximum rotation velocity in the cyclic-C₃H₂ emission is the position of the centrifugal barrier.

We simulate the PV diagram by using the radius of the centrifugal barrier and the maximum rotation velocity as parameters. In the simulation, we consider only the velocity field, and the highest and lowest velocities at a given offset from the protostar (Fig. 2b) are compared with the PV diagram of cyclic-C₃H₂. The best-fitting result is shown in Fig. 2c. The overall behaviour of the PV diagram is well explained even by this simple model. From this result, the radius of the centrifugal barrier and the maximum rotation velocity are derived to be 100 ± 20 AU and 1.8 ± 0.2 km s^{–1}, respectively. The radius of the centrifugal barrier is close to the radius of the Keplerian disk (90 AU)⁶. The total mass of the protostar and the inner disk is evaluated to be 0.18 ± 0.05 solar masses, by using the radius of the centrifugal barrier and the maximum rotation velocity derived above. Here, the error is estimated from those of the radius of the centrifugal barrier and the maximum rotation velocity by error propagation. This is consistent with the protostar mass reported in ref. 6 (0.19 ± 0.04 solar masses) from the measurement of the Keplerian rotation.

In contrast to the cyclic-C₃H₂ lines, the SO ($J_N = 7_8–6_7$) line shows a compact single-peaked distribution centred at the protostar (Fig. 1a) (J_N represents the rotational energy level of SO; see the footnote of Extended Data Table 1). The SO distribution seems to fill up the dip of the cyclic-C₃H₂ distribution. The PV diagram of the SO line emission along the north–south line is completely different from that of cyclic-C₃H₂ (Fig. 1d). It consists of two components. One is a component whose velocity is proportional to the position offset from the protostar. The other is a weak and broad component whose velocity width is as broad as 6 km s^{–1}. The anticorrelation between the distributions of cyclic-C₃H₂ and SO is, thus, evident. When we compare the PV diagram of SO with that of cyclic-C₃H₂, it is most likely that the former component of SO comes from a rotating ring whose radius is close to the innermost radius of the cyclic-C₃H₂ distribution, corresponding to the radius of

¹Department of Physics, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ²Department of Communication Engineering and Informatics, The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan. ³National Astronomical Observatory of Japan, Osawa, Mitaka, Tokyo 181-8588, Japan. ⁴Institut de Planétologie et d'Astrophysique de Grenoble, BP 53, 38041 Grenoble Cedex 9, France. ⁵Université de Toulouse, Université Paul Sabatier, Observatoire Midi-Pyrénées (UPS-OMP), Institut de Recherche en Astrophysique et Planétologie (IRAP), Toulouse, France. ⁶Centre National de la Recherche Scientifique (CNRS), IRAP, 9 Avenue Colonel Roche, BP 44346, Toulouse 31028 Cedex 4, France. ⁷Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, Copenhagen 2100 Østerbro, Denmark. ⁸Centre for Star and Planet Formation and Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, Copenhagen 1350 K, Denmark. ⁹National Aeronautics and Space Administration (NASA), Goddard Space Flight Center (GSFC), Astrochemistry Laboratory, Mail Code 691, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ¹⁰Academia Sinica Institute of Astronomy and Astrophysics, PO Box 23-141, Taipei 10617, Taiwan. ¹¹Subaru Telescope, National Astronomical Observatory of Japan, 650 North A'ohoku Place, Hilo, Hawaii 96720, USA. ¹²Institute of Astrophysics, National Taiwan University, Taipei 10617, Taiwan. ¹³Department of Earth and Planetary Sciences, Kobe University, Kobe 657-8501, Japan.

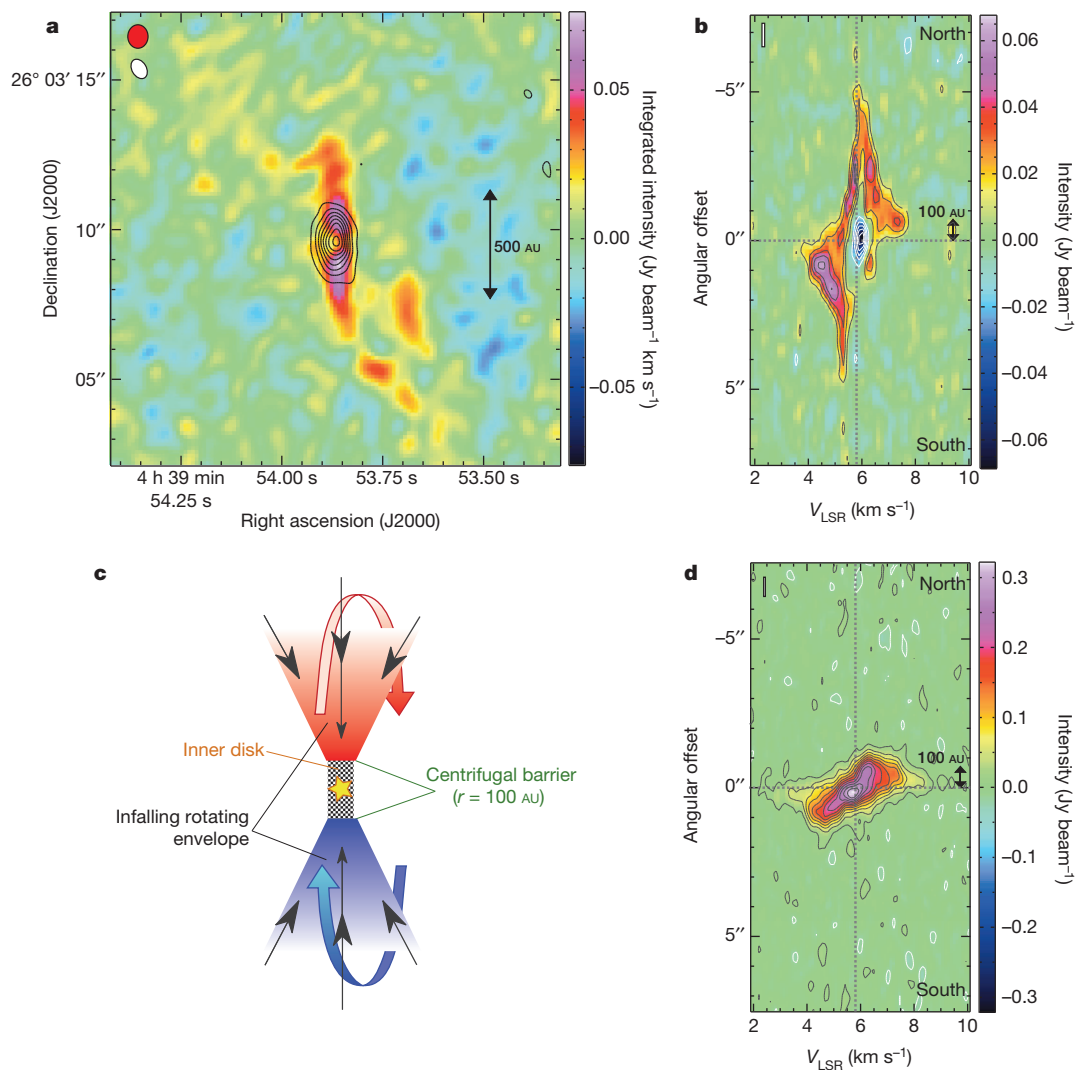


Figure 1 | IRAS 04368+2557 in the cyclic- C_3H_2 ($5_{23,432}$) and SO ($J_N = 7_6-6_5$) lines. **a**, Integrated intensity distributions of cyclic- C_3H_2 (colour) and SO (contours). Contours are every 10σ ($90 \text{ mJy beam}^{-1} \text{ km s}^{-1}$), starting from 5σ (the outermost contour). The red and white ellipses represent the synthesized beam sizes for cyclic- C_3H_2 and SO, respectively. **b**, The PV diagram of cyclic- C_3H_2 along a north-south line passing through the protostar position. Contours are every 3σ (12 mJy beam^{-1}), starting from 3σ . The white rectangle represents the resolution. **c**, A schematic of the envelope geometry. **d**, The PV diagram of SO along the same line as **b**. Contours are every 4σ (28 mJy beam^{-1}), starting from 2σ . White contours represent negative values. V_{LSR} denotes the velocity with respect to the local standard of rest. The white rectangle represents the resolution.

the centrifugal barrier (Fig. 1c). On the other hand, the latter component seems to come from the inner disk, although it is much fainter than the ring component. These features are essentially the same for the other SO line ($J_N = 7_6-6_5$). We conclude that a drastic chemical change occurs at the centrifugal barrier. (See the Methods and Extended Data Figs 2, 4, 5, 6 and 7 for details.)

We observed four lines of cyclic- C_3H_2 and two lines of SO in total (Extended Data Table 1). We analysed them to derive physical conditions of emitting regions for both molecules by using a non-local-thermodynamic-equilibrium (non-LTE) large-velocity-gradient (LVG) code (Methods and Extended Data Fig. 8). The gas kinetic temperatures of the cyclic- C_3H_2 emitting region are derived to be 23–33 K, 30 K and 23 K, for the $1'' \times 1''$ region at the centre, $1''$ north of the centre and $1''$ south of the centre, respectively, where the H_2 density ranges from 6×10^6 to $1 \times 10^8 \text{ cm}^{-3}$. These values are almost consistent with the reported model⁹ (H_2 density of $1.3 \times 10^8 \text{ cm}^{-3}$ and the temperature of 30 K at 100 AU). On the other hand, the gas kinetic temperature of the SO-emitting region is constrained to be higher than 60 K, and the H_2 density to be higher than $3 \times 10^6 \text{ cm}^{-3}$ for the $1'' \times 2''$ region centred at the protostar position.

Given that the sublimation temperature of SO is 50 K, it is most likely that SO is liberated into the gas phase from grain mantle at the centrifugal barrier. Although the iced SO is not identified by infrared observations, it is predicted by the chemical model^{10,11}. Liberation of the S atom into the gas phase followed by the reaction with OH may also contribute to enhancement of SO (ref. 12). Possible mechanisms for the SO liberation are accretion shock in front of the centrifugal barrier,

outflow shock on the disk surface, and the protostellar heating. These possibilities are described in the Supplementary Information.

Further in the mid-plane of the inner disk, SO is probably depleted again onto dust grains. The depletion timescale is as short as 10–200 years for the H_2 density of $6 \times 10^6 \text{ cm}^{-3}$ to reach $1 \times 10^8 \text{ cm}^{-3}$. Therefore, SO mainly exists around the centrifugal barrier. However, the PV diagram of the SO emission also shows a weak broad component inward of the centrifugal barrier. This suggests that SO partly survives even inward of the barrier. The gas infalling directly onto the inner disk surface may also contribute to the broad component¹³. We predict the velocity of the Keplerian rotation as a function of the radius from the estimated protostellar mass derived from the cyclic- C_3H_2 data (Fig. 2c). The expected velocity can roughly explain the PV diagram inward of the centrifugal barrier.

Similarly, cyclic- C_3H_2 is depleted onto dust grains in the mid-plane of the disk as in the case of SO. Cyclic- C_3H_2 may also be destroyed in the transition region by a gas-phase reaction with the oxygen atom. Although the fresh gas containing cyclic- C_3H_2 is continuously supplied outside the centrifugal barrier, it is not supplied efficiently within the disk. Hence, the abundance of cyclic- C_3H_2 in the gas phase becomes much lower inward of the centrifugal barrier than in the infalling rotating envelope, as observed.

These results demonstrate a drastic chemical change at the centrifugal barrier in the course of the formation of the inner disk. It does not mean that SO and cyclic- C_3H_2 are chemically related to each other directly; the change is caused by the discontinuous infalling motion of the gas at

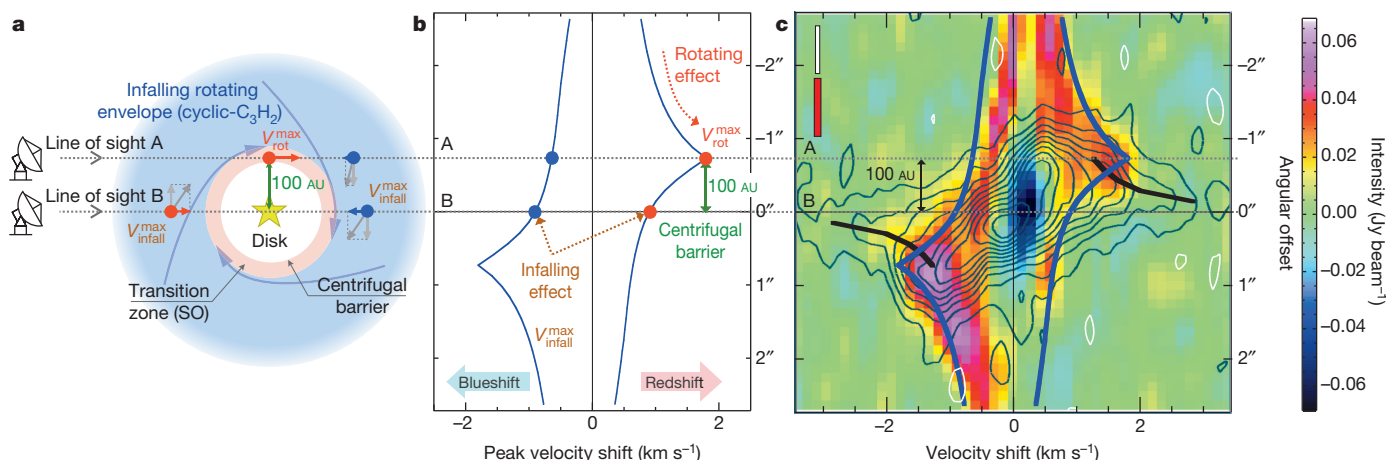


Figure 2 | A model of an infalling rotating envelope. **a**, A schematic illustration of the infalling rotating envelope. The observer is on the left-hand side, looking at the envelope in an edge-on configuration. **b**, The highest and lowest velocities calculated with the toy model (Methods). The emissions of the coloured closed circles come from the corresponding coloured closed circles

the centrifugal barrier and local heating processes there (Supplementary Information). When interstellar matter is brought into the disk, it must experience this physical situation and is subject to significant chemical processing including gas–grain interactions in the transition zone. The chemical compositions of interstellar clouds have arbitrarily been assumed as the initial conditions set for the chemical evolution models of the disk^{14–17}. This assumption has now been found to be over-simplified: the chemical processing in the transition zone should be seriously considered. This is also true for the case of tracing the origin of the pre-solar materials found in meteorites back to interstellar matter.

METHODS SUMMARY

Observations were carried out with the Atacama Large Millimeter/submillimetre Array (ALMA) in 2012. We observed the rotational spectral line emissions of cyclic- C_3H_2 and SO as well as the dust continuum emission in the 249-GHz, 261-GHz, 338-GHz and 351-GHz regions. The synthesized beam is about $0.8'' \times 0.7''$ for 249 GHz and 261 GHz and about $0.7'' \times 0.5''$ for 338 GHz and 351 GHz. The integrated intensity maps and the PV diagrams of the four cyclic- C_3H_2 lines and the two SO lines are similar to each other for each molecule. For SO, the total flux of the $J_N = 7_8-6_7$ line emission observed with ALMA recovers 60% of that observed with the single-dish telescope in the Atacama Submillimetre Telescope Experiment (ASTE) (with a beam size of $22''$), indicating that the SO line emission is concentrated around the protostar. The effect of the outflow is confirmed to be negligible by examining the PV diagram of the SO ($J_N = 7_8-6_7$) line along the outflow axis. To account for the PV diagrams of the C_3H_2 lines, we use a simple kinematic model for the infalling rotating envelope, revealing that the maximum infalling velocity towards the protostar is just half of the maximum rotation velocity. We evaluated the gas kinetic temperature and the H_2 density using the non-LTE LVG code.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 July; accepted 19 December 2013.

Published online 12 February 2014.

1. Torres, R. M., Loinard, L., Mioduszewski, A. J. & Rodriguez, L. F. VLBA determination of the distance to nearby star-forming regions. II. Hubble 4 and HDE 283572 in Taurus. *Astrophys. J.* **671**, 1813–1819 (2007).
2. Andre, P., Ward-Thompson, D. & Barsony, M. Submillimeter continuum observations of Rho Ophiuchi A—the candidate protostar VLA 1623 and prestellar clumps. *Astrophys. J.* **406**, 122–141 (1993).
3. Ohashi, N., Hayashi, M., Ho, P. T. P. & Momose, M. Interferometric imaging of IRAS04368+2557 in the L1527 molecular cloud core: a dynamically infalling envelope with rotation. *Astrophys. J.* **475**, 211–223 (1997).
4. Shirley, Y. L., Evans, N. J. II & Rawlings, J. M. C. Tracing the mass during low-mass star formation. III. Models of the submillimeter dust continuum emission from class 0 protostars. *Astrophys. J.* **575**, 337–353 (2002).

in **a**, **c**. Here the highest and lowest velocities shown in **b** (thick blue lines) are superposed on the PV diagram of cyclic- C_3H_2 . The thick black lines indicate the Keplerian velocity expected for the disk. They trace the broad line profile of SO (thin grey contours) well. White and red rectangles represent the resolutions for SO and cyclic- C_3H_2 , respectively.

5. Yen, H.-W., Takakuwa, S., Ohashi, N. & Ho, P. T. P. Unveiling the evolutionary sequence from infalling envelopes to Keplerian disks around low-mass protostars. *Astrophys. J.* **772**, 22 (2013).
6. Tobin, J. J. *et al.* A ~ 0.2 -solar-mass protostar with a Keplerian disk in the very young L1527 IRS system. *Nature* **492**, 83–85 (2012).
7. McKee, C. F. & Ostriker, E. C. Theory of star formation. *Annu. Rev. Astron. Astrophys.* **45**, 565–687 (2007).
8. Williams, J. P. & Cieza, L. A. Protoplanetary disks and their evolution. *Annu. Rev. Astron. Astrophys.* **49**, 67–117 (2011).
9. Tobin, J. J. *et al.* Modeling the resolved disk around the class 0 protostar L1527. *Astrophys. J.* **771**, 48 (2013).
10. Aikawa, Y., Wakelam, V., Garrod, R. T. & Herbst, E. Molecular evolution and star formation: from prestellar cores to protostellar cores. *Astrophys. J.* **674**, 993–1005 (2008).
11. Aikawa, Y., Wakelam, V., Hersant, F., Garrod, R. T. & Herbst, E. From prestellar to protostellar cores. II. Time dependence and deuterium fractionation. *Astrophys. J.* **760**, 40 (2012).
12. Wakelam, V. *et al.* Sulphur chemistry and molecular shocks: the case of NGC1333-IRAS2. *Astron. Astrophys.* **437**, 149–158 (2005).
13. Neufeld, D. A. & Hollenbach, D. J. Dense molecular shocks and accretion onto protostellar disks. *Astrophys. J.* **428**, 170–185 (1994).
14. Herbst, E. & van Dishoeck, E. F. Complex organic interstellar molecules. *Annu. Rev. Astron. Astrophys.* **47**, 427–480 (2009).
15. Caselli, P. & Ceccarelli, C. Our astrochemical heritage. *Astron. Astrophys. Rev.* **20**, 56 (2012).
16. Visser, R., Doty, S. D. & van Dishoeck, E. F. The chemical history of molecules in circumstellar disks II. Gas-phase species. *Astron. Astrophys.* **534**, A132 (2011).
17. Hincelin, U., Wakelam, V., Commerçon, B., Hersant, F. & Guilloteau, S. Survival of interstellar molecules to prestellar dense core collapse and early phases of disk formation. *Astrophys. J.* **775**, 44 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Hanawa and K. Furuya for discussions. This paper makes use of the ALMA dataset ADS/JAO.ALMA#2011.0.00604.S. ALMA is a partnership of the ESO (representing its member states), the NSF (USA) and NINS (Japan), together with the NRC (Canada) and the NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by the ESO, the AUI/NRAO and the NAOJ. We thank the ALMA staff for their support. N.S. and S.Y. acknowledge financial support from Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technologies of Japan (21224002, 25400223 and 25108005), and by JSPS and MAEE under the Japan–France integrated action programme (SAKURA). T.H. acknowledges financial support from Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technologies of Japan (21224002, 24684011 and 25108005). C.C. and C.K. acknowledge financial support from the French Agence Nationale pour la Recherche (ANR) project FORCOMS (contract ANR-08-BLAN-0225) and from the Partenariats Hubert Curien (PHC) Programme SAKURA 25765VC.

Author Contributions N.S. led the project and participated in data reduction. All authors contributed to the data analysis, discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.S. (nam@taurus.phys.s.u-tokyo.ac.jp).

Optical detection of radio waves through a nanomechanical transducer

T. Bagci¹, A. Simonsen¹, S. Schmid², L. G. Villanueva², E. Zeuthen¹, J. Appel¹, J. M. Taylor³, A. Sørensen¹, K. Usami¹, A. Schliesser¹ & E. S. Polzik¹

Low-loss transmission and sensitive recovery of weak radio-frequency and microwave signals is a ubiquitous challenge, crucial in radio astronomy, medical imaging, navigation, and classical and quantum communication. Efficient up-conversion of radio-frequency signals to an optical carrier would enable their transmission through optical fibres instead of through copper wires, drastically reducing losses, and would give access to the set of established quantum optical techniques that are routinely used in quantum-limited signal detection. Research in cavity optomechanics^{1,2} has shown that nanomechanical oscillators can couple strongly to either microwave^{3–5} or optical fields^{6,7}. Here we demonstrate a room-temperature optoelectromechanical transducer with both these functionalities, following a recent proposal⁸ using a high-quality nanomembrane. A voltage bias of less than 10 V is sufficient to induce strong coupling^{4,6,7} between the voltage fluctuations in a radio-frequency resonance circuit and the membrane's displacement, which is simultaneously coupled to light reflected off its surface. The radio-frequency signals are detected as an optical phase shift with quantum-limited sensitivity. The corresponding half-wave voltage is in the microvolt range, orders of magnitude less than that of standard optical modulators. The noise of the transducer—beyond the measured 800 pV Hz^{−1/2} Johnson noise of the resonant circuit—consists of the quantum noise of light and thermal fluctuations of the membrane, dominating the noise floor in potential applications in radio astronomy and nuclear magnetic imaging. Each of these contributions is inferred to be 60 pV Hz^{−1/2} when balanced by choosing an electromechanical cooperativity of ~ 150 with an optical power of 1 mW. The noise temperature of the membrane is 300 K divided by the cooperativity. For the highest observed cooperativity of 6,800, this leads to a projected noise temperature of 40 mK and a sensitivity limit of 5 pV Hz^{−1/2}. Our approach to all-optical, ultralow-noise detection of classical electronic signals sets the stage for coherent up-conversion of low-frequency quantum signals to the optical domain^{8–11}.

Optomechanical and electromechanical systems^{1,2} have gained considerable attention recently for their potential as hybrid transducers between otherwise incompatible (quantum) systems, such as photonic, electronic and spin degrees of freedom^{2,10,12}. The mechanical coupling of radio-frequency or microwave signals to optical fields is particularly attractive for present-day and future quantum technologies. Photon-phonon transfer protocols viable all the way to the quantum regime have already been implemented separately in the radio- and optical-frequency domains^{7,13,14}.

Among the optomechanical systems that have been considered for radio-to-optical transduction^{8–10,15}, we choose an approach⁸ based on a nanomembrane^{16,17} with a high quality factor, $Q_m \approx 3 \times 10^5$, which is coupled capacitively¹⁸ to a radio-frequency resonance circuit (Fig. 1). Together with a four-segment gold electrode, the membrane forms a capacitor, whose capacitance, $C_m(x)$, depends on the membrane-electrode distance, $d + x$. With a tuning capacitor C_0 , the total capacitance, $C(x) = C_0 + C_m(x)$, forms a resonance circuit with a typical

quality factor $Q_{LC} = \sqrt{L/C}/R = 130$ using a custom-made coil wired on a low-loss ferrite rod. This yields an inductance $L = 0.64$ mH and a loss $R \approx 20 \Omega$. The circuit's resonance frequency $\Omega_{LC} = 1/\sqrt{LC}$ is tuned to the frequency, $\Omega_m/2\pi = 0.72$ MHz, of the fundamental drum mode of the membrane. The membrane-circuit system is coupled to a propagating optical mode reflected from the membrane.

The electromechanical dynamics is described most generically by the Hamiltonian⁸

$$H = \frac{\phi^2}{2L} + \frac{p^2}{2m} + \frac{m\Omega_m^2 x^2}{2} + \frac{q^2}{2C(x)} - qV_{dc} \quad (1)$$

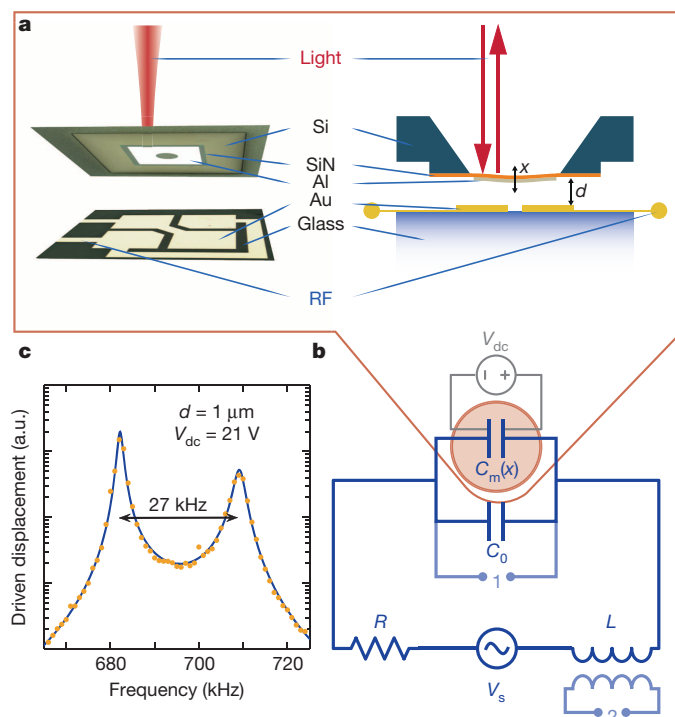


Figure 1 | Optoelectromechanical system. **a**, A 500- μm -square membrane of Al-coated³⁰ SiN in vacuum ($<10^{-5}$ mbar) forms a position-dependent capacitor, $C_m(x=0) \approx 0.5$ pF, with a planar, four-segment gold electrode in the immediate vicinity ($0.9 \mu\text{m} \lesssim d \lesssim 6 \mu\text{m}$). The membrane electrode's potential is electrically floating. The membrane's displacement is converted into a phase shift of the laser beam reflected from the membrane. **b**, The membrane capacitor is part of an LC circuit, tuned to the mechanical resonance frequency by means of a tuning capacitor $C_0 \approx 80$ pF (Supplementary Information). A bias voltage, V_{dc} , couples the excitations of the LC circuit to the membrane's motion. The circuit is driven by a voltage V_s , which can be injected through the coupling port 2 or picked up by the inductor from the ambient radio-frequency (RF) radiation. **c**, For $d = 1 \mu\text{m}$, the optically observed response of the membrane to a weak excitation of the system shows a split peak due to strong electromechanical coupling. a.u., arbitrary units.

¹Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark. ²Department of Micro- and Nanotechnology, Technical University of Denmark, DTU Nanotech, 2800 Kongens Lyngby, Denmark. ³Joint Quantum Institute/NIST, College Park, Maryland 20742, USA.

where ϕ and q , respectively the flux in the inductor and the charge on the capacitors, are conjugate variables for the LC circuit, and x and p respectively denote the position and momentum of the membrane with an effective mass m . The last two terms represent the charging energy, $U_C(x)$, of the capacitors, which can be offset using an externally applied bias voltage, V_{dc} (Fig. 1). This energy, corresponding to the charge $\bar{q} = V_{dc}C(\bar{x})$, leads to a new equilibrium position, \bar{x} , for the membrane. Furthermore, the position-dependent capacitive force $F_C(x) = -dU_C/dx$ causes spring softening, reducing the membrane's motional eigenfrequency by $\Delta\Omega_m \approx -C'(\bar{x})V_{dc}^2/2m\Omega_m$ (ref. 19).

Much richer dynamics than this shift may be expected from the mutually coupled system described by equation (1). For small excursions, $(\delta q, \delta x)$, around the equilibrium, (\bar{q}, \bar{x}) , it can be described by the linearized interaction term⁸ (Supplementary Information)

$$H_I = G\delta q\delta x = h_{em} \frac{\delta q}{\sqrt{\hbar/2L\Omega_{LC}}} \frac{\delta x}{\sqrt{\hbar/2m\Omega_m}} \quad (2)$$

parameterized by either the coupling parameter $G = -V_{dc}C'(\bar{x})/C(\bar{x})$ or the electromechanical coupling energy, h_{em} , where \hbar is Planck's constant (\hbar) divided by 2π . This coupling leads to an exchange of energy between the electronic and mechanical subsystems at the rate g_{em} ; if this rate exceeds their dissipation rates, respectively $\Gamma_{LC} = \Omega_{LC}/Q_{LC}$ and $\Gamma_m = \Omega_m/Q_m$, they hybridize into a strongly coupled electromechanical system^{4,6,7}. Our system is deeply in the strong coupling regime ($2g_{em} = 2\pi \times 27 \text{ kHz} > \Gamma_{LC} = 2\pi \times 5.5 \text{ kHz} \gg \Gamma_m = 2\pi \times 20 \text{ Hz}$) for a distance $d = 1 \mu\text{m}$ and a bias voltage $V_{dc} = 21 \text{ V}$ (Fig. 1c). Here we detect the strong coupling using an independent optical probe on the mechanical system.

We have performed a series of experiments in which the bias voltage is systematically increased with a different sample, a larger distance, $d = 5.5 \mu\text{m}$, and a lower mechanical dissipation, $\Gamma_m/2\pi = 2.3 \text{ Hz}$. The system is excited inductively through port 2 (Fig. 1b), inducing a weak radio-frequency signal of root mean squared amplitude $V_s = 670 \text{ nV}$, at a frequency $\Omega \approx \Omega_{LC}$. The response of the coupled system can be measured both electrically, as the voltage across the capacitors (port 1 in Fig. 1b), and optically, by analysing the phase shift of a light beam (wavelength $\lambda = 633 \text{ nm}$) reflected from the membrane. Both signals are recorded with a lock-in amplifier, which also provides the excitation signal.

The electrically measured response (Fig. 2a) shows the signature of a mechanically induced transparency²⁰, indicated by the dip in the LC resonance curve. Independently, we observe the radio-frequency signal in the LC circuit optically via the membrane mechanical dynamics (Fig. 2b). In particular, the electromechanical coupling leads to a broadening of the mechanical resonance to a new effective linewidth, $\Gamma_{eff} = (1 + C_{em})\Gamma_m$, where C_{em} is the electromechanical cooperativity:

$$C_{em} = \frac{4g_{em}^2}{\Gamma_m\Gamma_{LC}} \quad (3)$$

The width of the induced transparency dip and the mechanical linewidth grow in unison, and in agreement with our expectations, as $\Gamma_{eff} \propto V_{dc}^2$ (Fig. 2b, inset). Both of these features also shift to lower frequencies as the bias voltage is increased, following the expected $\Delta\Omega_m \propto -V_{dc}^2$ dependence¹⁹. We note that in each experiment we have tuned the LC resonance frequency to Ω_m .

Using the model based on the full Langevin equations (Supplementary Information), derived from the Hamiltonian in equation (1), we fit the electronically and optically measured curves, and for the two curves obtain fit parameters Ω_m , Ω_{LC} , Γ_{LC} and G that agree typically to within 1%. Together with the intrinsic damping determined independently from thermally driven spectra, the system's dynamics can be quantitatively predicted. Our data analysis allows us to quantify the coupling strength in three independent ways: analysis of the mechanical responses' spectral shape; comparison of the voltage and displacement modulation amplitudes; and in terms of the frequency

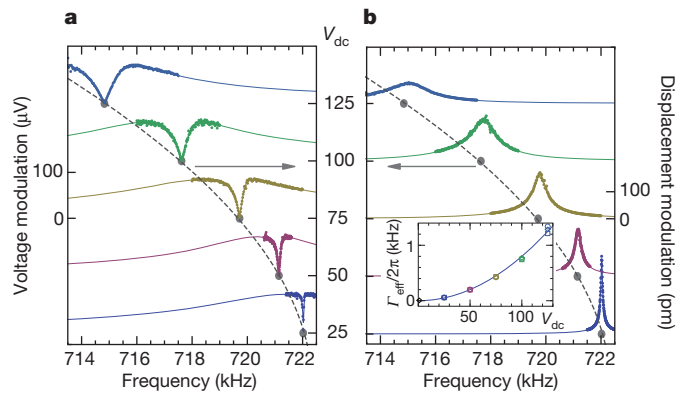


Figure 2 | Mechanically induced transparency. Response of the coupled system to a weak excitation at frequency Ω (through port 2 in Fig. 1b) probed through the voltage modulation in the LC circuit (at port 1) (a) and through the optical phase shift (b). The data (coloured dots) measured for five different bias voltages agree excellently with model fits (curves) respectively corresponding to $g_{em}/2\pi = 280, 470, 810, 1,030$ and $1,290 \text{ Hz}$ (from bottom to top). Each curve is offset so that its baseline corresponds to the V_{dc} indicated between the panels. Grey points indicate Ω_m values extracted for each set of data. A shift of $\Delta\Omega_m \propto -V_{dc}^2$ is fitted with the dashed line. Inset, effective linewidth of the mechanical resonance extracted from full model fits to the electrically (circles) and optically (boxes) measured response and simple Lorentzian fits to the optical data (diamonds). The solid line shows the expected $\Gamma_{eff} \propto V_{dc}^2$ scaling.

shift¹⁹ of the mechanical mode. Finally, we compare these experimental values with a theoretical estimate accounting for the geometry of the electromechanical transducer. For $V_{dc} = 125 \text{ V}$, we find that $G = 10.3 \text{ kV m}^{-1}$ following the first method, and similar values using the three others (Supplementary Information), demonstrating our thorough understanding of the system.

In another experimental run ($d = 4.5 \mu\text{m}$; Fig. 3), we characterized the strong electromechanical coupling^{3,4,14} using the normal-mode splitting that gives rise to an avoided crossing of the resonances of the electronic circuit and the mechanical mode, as the latter is tuned through the former using the capacitive spring effect¹⁹. In contrast to earlier observations^{4,6,7}, we simultaneously witness the strong coupling through the optical readout, in which the recorded light phase reproduces the membrane motion (Fig. 3c, e). Again, the predictions derived from the Langevin equations are in excellent agreement with our observations, yielding a cooperativity of $C_{em} = 3,800$ for these data with $m = 24 \text{ ng}$ and $\Gamma_m/2\pi = 3.1 \text{ Hz}$.

We now turn to the performance of this interface as a radio-frequency/optical transducer. A relevant figure of merit for the purpose of bringing small signals onto an optical carrier is the voltage, V_π , required at the input of the series circuit to induce an optical phase shift of π . Achieving minimal V_π requires a balance between strong coupling and induced mechanical damping. For the optimal cooperativity, $C_{em} = 1$, we find that

$$V_\pi = \frac{1}{2} \sqrt{mL\Gamma_m\Gamma_{LC}\lambda\Omega_r} \approx 140 \mu\text{V} \quad (4)$$

at resonance ($\Omega_r \equiv \Omega_m = \Omega_{LC} = \Omega$), which is orders of magnitude below the corresponding figure of merit for not only commercial modulators optimized for decades by the telecom industry, but also explorative microwave photonic devices^{21,22} based on electronic nonlinearities. It is interesting to relate this performance to more fundamental entities, namely the electromagnetic field quanta that constitute the signal. Indeed it is possible to show that the quantum conversion efficiency, defined here as the ratio of optical sideband photons to the radio-frequency quanta extracted from the source, $V_s I / \hbar \Omega_{LC}$, for $C_{em} \gg 1$, is given by (Supplementary Information)

$$\eta_{eo} = 4(kx_{zpf})^2 \frac{\Phi_{car}}{\Gamma_m} \quad (5)$$

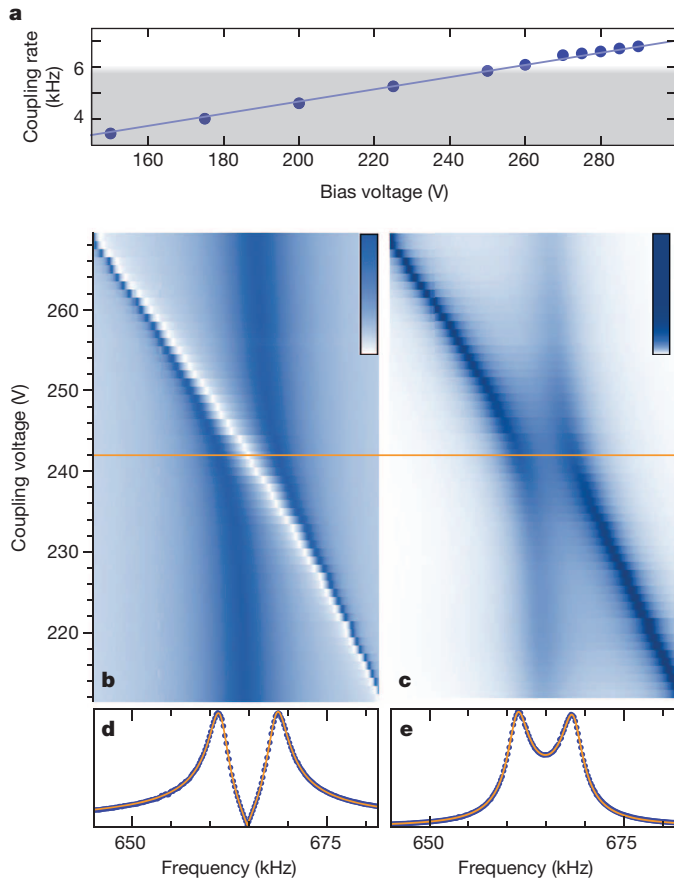


Figure 3 | Strong-coupling regime. **a**, Measured coherent coupling rate, $2g_{em}/2\pi$, as a function of bias voltage (points), and linear fit (line). The shaded area indicates the dissipation rate, $\Gamma_{LC}/2\pi \approx 5.9$ kHz of the LC circuit. **b–e**, Normalized response of the coupled system as measured on port 1 (Fig. 1c) (**b**, **d**) and via the optical phase shift induced by membrane displacements (**c**, **e**). The colour scales encode normalized voltage (**b**) and displacement modulation (**c**). On tuning of the bias voltage, the mechanical resonance frequency is tuned through the LC resonance, but owing to the strong coupling an avoided crossing is very clearly observed. Panels **d** and **e** show the spectra corresponding to the orange lines in **b** and **c**, at $V_{dc} = 242$ V, where the electronic and mechanical resonance frequencies coincide. Points are data; orange line is the model fit.

This corresponds to the squared effective Lamb–Dicke parameter, $(kx_{zpt})^2 = (2\pi/\lambda)^2 \hbar/2m\Omega_m$, enhanced by the number of photons sampling the membrane during the membrane excitations’ lifetime (Φ_{car} is the photon flux and k is the wavenumber). For the experiments shown in Fig. 2, we deduce a conversion efficiency of 0.8% from the independently measured radio-frequency voltage and optical phase modulation. Although this result is limited by the optical power in this interferometer, we performed tests to confirm that the membranes can support optical readout powers of more than $\Phi_{car}\hbar c/\lambda = 20$ mW without degradation of their (intrinsic) linewidth. We thus project that conversion efficiencies of the order of 50% are available. Note that this transducer constitutes a phase-insensitive amplifier, and can thus reach conversion efficiencies greater than one—at the expense of added quantum noise.

For the recovery of weak signals, the sensitivity and bandwidth of the interface is of greatest interest. The signal at the optical output of the device is the interferometrically measured spectral density of the optical phase, ϕ , of the light reflected from the membrane:

$$S_{\phi\phi}^{\text{tot}} = (2k)^2 |\chi_m^{\text{eff}}|^2 (|G\chi_{LC}|^2 S_{VV} + S_{FF}^{\text{th}}) + S_{\phi\phi}^{\text{im}} \quad (6)$$

The voltage V_s at the input of the resonance circuit (denoted here as its spectral density S_{VV}) is transduced to a phase shift via the circuit’s susceptibility, χ_{LC} , the coupling, G , the effective membrane susceptibility,

χ_m^{eff} , and the optical wavenumber, k (Supplementary Information). The sensitivity is determined by the noise added within the interface. This includes, in particular, the imprecision in the phase measurement ($S_{\phi\phi}^{\text{im}}$) and the random thermal motion of the membrane induced by the Langevin force (S_{FF}^{th}). The former depends on the performance of the interferometric detector used and can be quantum limited ($S_{\phi\phi}^{\text{im}} \propto \Phi_{car}^{-1}$).

We demonstrate the sensitivity and the noise performance of the transduction scheme by measuring the noise as a function of the input circuit resistor and its temperature (Fig. 4). Because the home-made, high- Q inductor is too sensitive to the ambient radio-frequency radiation (Supplementary Information), we use a shielded commercial inductor (Picoelectronics) resulting in a lower value, $Q_{LC} = 47$, for these measurements. Red traces in Fig. 4a and Fig. 4b respectively present the optically measured noise spectrum and the corresponding voltage noise. On resonance, the dominant contribution is the Johnson

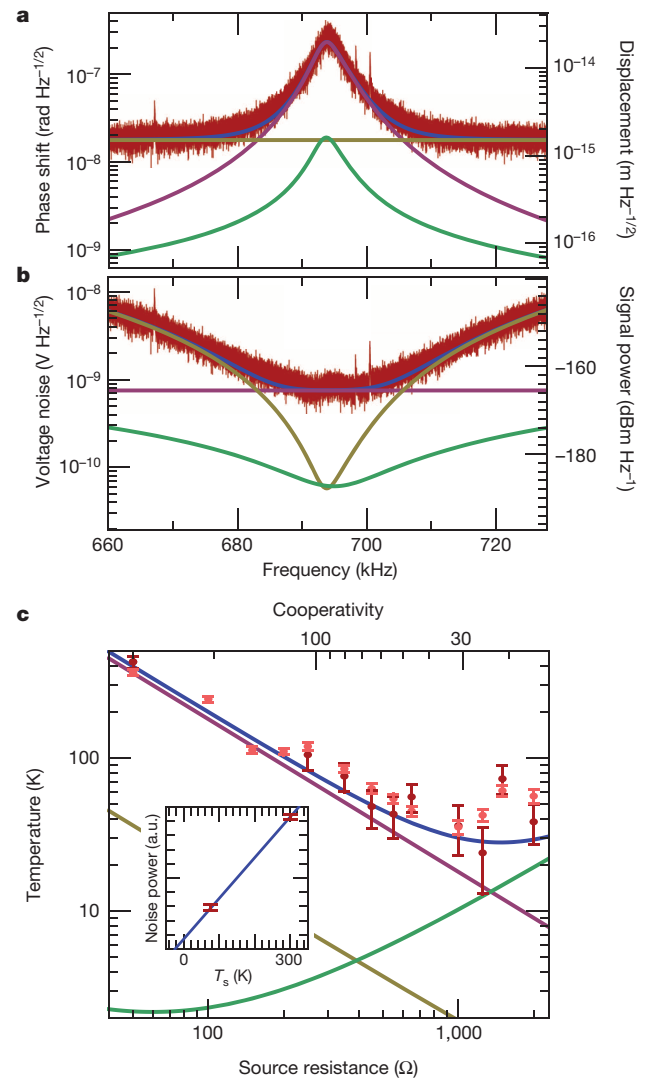


Figure 4 | Voltage sensitivity and noise. Noise characterization of the transducer with contributions from Johnson noise (violet), optical quantum phase noise (yellow) and membrane thermal noise (green). **a**, Optically measured noise (red) is well reproduced by a model $\sqrt{S_{\phi\phi}^{\text{tot}}}$ (blue). **b**, Data and models as in **a**, but divided by the interface’s response function, $|\chi_{tot}|$, and thus referenced to the voltage, V_s , induced in the circuit. **c**, Noise temperature of the amplifier (errors, s.d.). It is determined using the Y-factor method, at the resonance frequency (dark red points), and in a 10-kHz-wide band around the resonance (light red points), as a function of external loading. Lines are the model of equation (7), broken down into contributions as in **a** and **b**. Inset, example of a noise temperature measurement at $R_s = 1,250$ Ω.

noise ($S_{VV}^j \approx (800 \text{ pV})^2 \text{ Hz}^{-1}$) of the circuit (violet). Off resonance, optical quantum (shot) noise (yellow) limits the phase sensitivity to $S_{\phi\phi}^{\text{sim}} = (18 \text{ nrad})^2 \text{ Hz}^{-1}$, corresponding to membrane displacements of $(1.5 \text{ fm})^2 \text{ Hz}^{-1}$. In this experiment, we used a home-built interferometer operating at $\lambda = 1,064 \text{ nm}$ and with a light power of $\sim 1 \text{ mW}$ returned from a membrane with $m = 64 \text{ ng}$ and $\Gamma_m/2\pi = 20 \text{ Hz}$. The square root of the phase sensitivity, $\sqrt{S_{\phi\phi}^{\text{sim}}}$, can be translated into a voltage sensitivity limit by division by the transfer function $|\chi^{\text{tot}}| \equiv |2k\chi_m^{\text{eff}}G\chi_{\text{LC}}|$ of the transducer. With the cooperativity chosen here, $C_{\text{em}} = 150$, this corresponds to a voltage noise level of $60 \text{ pV Hz}^{-1/2}$ within the resonant bandwidth of this proof-of-principle transducer, but higher powers, and more sensitive optomechanical transduction^{16,23} could readily improve this number. From our model, we furthermore deduce that the contribution of the thermal motion of the membrane adds an equal amount of voltage noise (Fig. 4b, green), such that, at this cooperativity, these noise contributions are balanced and their sum minimized to $84 \text{ pV Hz}^{-1/2}$.

Further analysis of the transducer noise has been performed by measurements with an additional, ‘source’, resistor, R_s , in series with the inductor of the circuit (Fig. 4c). The input to the circuit thus consists of the Johnson noise of both resistors, $(S_{VV}^j)' = 2k_B(RT_R + R_sT_s)$. We cool the source resistor using liquid nitrogen, and optically measure the displacement of the membrane both at room ($T_s = 300 \text{ K}$) and at liquid nitrogen temperature ($T_s = 77 \text{ K}$). We can thus determine the amount of noise added by the transducer using the Y-factor method²⁴. From equation (6), we expect to find a noise temperature of (Supplementary Information)

$$T_n = \left(\frac{1}{\eta_e} - 1 \right) T_R + \frac{1}{\eta_e} \left(\frac{1}{C_{\text{em}}} T_m + \frac{(1 + C_{\text{em}})^2}{C_{\text{em}}} T_L \right) \quad (7)$$

at resonance, where the three summands are due to the Johnson noise of the circuit’s loss, $R = 60 \Omega$, at $T_R = 300 \text{ K}$, the membrane’s thermal fluctuations ($T_m = 300 \text{ K}$) and the noise in the optical readout ($T_L \approx 50 \text{ mK}$), respectively. We note that both the circuit’s loading, $\eta_e = R_s/(R_s + R)$, and the cooperativity, $C_{\text{em}} = RC_{\text{em}}(R_s = 0 \Omega)/(R_s + R)$ are now functions of the source resistance. In this experiment, we vary the cooperativity by varying the source resistor from $C_{\text{em}}(R_s = 0 \Omega) = 550$ to $C_{\text{em}}(R_s = 2 \text{ k}\Omega) = 18$, and find a noise temperature consistent overall with equation (7), with the lowest measured value reaching down to 24 K (Fig. 4c).

The challenge of engineering a low-loss, overcoupled electronic resonance circuit ($\eta_e \rightarrow 1$) aside, the transducer itself adds only very little noise (green and yellow lines in Fig. 4c, representing the second and third terms in equation (7), respectively). For example, at a cooperativity of $C_{\text{em}} = 70$ achieved with $R_s = 400 \Omega$, subtracting the Johnson noise from the total noise yields optical quantum phase noise and membrane thermal noise temperatures of 4 K . Remarkably, the membrane contribution, which can usually only be suppressed by cryogenic cooling, is strongly reduced by the cooperativity parameter ($\propto T_m/C_{\text{em}}$). The highest cooperativity we have obtained is $C_{\text{em}} = 6,800$, by applying equation (3) to the data of Fig. 1c. This implies that membrane noise temperatures down to 40 mK can be expected, corresponding here to a voltage noise level of $5 \text{ pV Hz}^{-1/2}$.

For comparison, we performed measurements with an arrangement of ultralow-noise operational amplifiers connected directly to port 1. The amplifier is based on junction field-effect transistors (JFETs) and combines low input voltage noise (nominally $4 \text{ nV Hz}^{-1/2}$) with extremely low current noise (nominally $2.5 \text{ fA Hz}^{-1/2}$), as required²⁴ for measurements on a relatively high source impedance, which here amounts to $RQ_{\text{LC}}^2 \approx 140 \text{ k}\Omega$ at port 1. In practice, with a gain of 1,000, the best voltage sensitivity we have obtained is only $S_{VV}^{\text{a,tot}}(\Omega_{\text{LC}}) = (130 \text{ pV})^2 \text{ Hz}^{-1}$ over the bandwidth of the LC resonance. Similar performance levels—on a par with the transducer discussed here—are expected even for ideal operation of other amplifiers described in the scientific and technical literature (Supplementary Information). Apart

from being competitive with standard electronics in its noise figures, our transducer provides a new functionality owing to its direct compatibility with fibre optical communication lines. The presented optoelectromechanical transducer also compares very favourably with previous proof-of-principle mechanical amplifiers for radio-frequency²⁵ and microwave²⁶ radiation (Supplementary Information).

Because our transducer noise floor is well below the room-temperature Johnson noise from the circuit’s loss, $R = 60 \Omega$, this approach can be of particular relevance in applications where electronic Johnson noise is suppressed. For example, for direct electronic (quantum) signal transduction, the resonance circuit is overloaded ($\eta_e \rightarrow 1$) with a cold transmission line that carries the signal of interest, but no Johnson noise. In radio astronomy²⁷, highly efficient antennas looking at the cold sky can have noise temperatures far below room temperature. The usually required cryogenically cooled pre-amplifiers might be replaced by our transducer—a critical advantage for satellite missions—and extension to gigahertz frequencies should be straightforward using a.c. driving⁴. Direct and efficient conversion of radio-frequency signals into optics could save substantial resources in large phased-array antennas. Finally, in NMR experiments including imaging, cooled pickup circuits can deliver a significant sensitivity improvement, but this approach challenges present amplifier technology^{28,29}.

METHODS SUMMARY

The capacitor is fabricated by standard clean-room microfabrication techniques. Electrodes made of gold (200 nm thick) are deposited on a glass substrate and structured by ion-beam etching. Each segment is $400 \mu\text{m}$ long, with $60\text{-}\mu\text{m}$ gaps between the segments. Pillars ranging in height from 600 nm to $1 \mu\text{m}$ are placed to define the membrane–electrode distance. The inductor is wound with Litz wires to ensure a high Q-factor. A variable trimming capacitor is used to tune the resonance frequency of the LC circuit.

The mechanical resonator consists of a 50-nm-thick aluminium layer on top of a high-stress stoichiometric SiN layer with a thickness of 100 or 180 nm depending on the sample. The aluminium layer is deposited on top of the whole wafer after the membranes have been released. Photolithography and chemical etching are subsequently used to remove the metal from the anchoring regions and from a circle in the middle of the membrane. The metal layer on SiN typically causes a 10% decrease in the eigenfrequency of the fundamental mode.

Optical interferometry is carried out using a commercial Doppler vibrometer (MSA-500 Polytec) and a home-made Michelson interferometer (for the data set in Fig. 4). The home-made Michelson interferometer uses shot-noise-limited, balanced-homodyne detection with a high-bandwidth (0–75 MHz) InGaAs receiver. The two d.c. outputs from the detector are used to generate the differential error signal, which is then fed to the piezoelectric actuator in the reference arm for locking the interferometer. The radio-frequency output of the detector is high-pass-filtered and fed to a spectrum analyser to record the vibrations of the membrane. Absolute calibration of the mechanical amplitude is carried out through a known modulation of the piezoelectric actuator at a frequency close to the mechanical peak.

Received 12 July 2013; accepted 13 January 2014.

- Kippenberg, T. J. & Vahala, K. J. Cavity optomechanics: back-action at the mesoscale. *Science* **321**, 1172–1176 (2008).
- Aspelmeyer, M., Kippenberg, T. J. & Marquardt, F. Cavity optomechanics. Preprint at <http://arxiv.org/abs/1303.0733> (2013).
- O’Connell, A. D. *et al.* Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
- Teufel, J. D. *et al.* Circuit cavity electromechanics in the strong-coupling regime. *Nature* **471**, 204–208 (2011).
- Faust, T., Krenn, P., Manus, S., Kotthaus, J. & Weig, E. Microwave cavity-enhanced transduction for plug and play nanomechanics at room temperature. *Nature Commun.* **3**, 728 (2012).
- Gröblacher, S., Hammerer, K., Vanner, M. R. & Aspelmeyer, M. Observation of strong coupling between a micromechanical resonator and an optical cavity field. *Nature* **460**, 724–727 (2009).
- Verhagen, E., Deleglise, S., Weis, S., Schliesser, A. & Kippenberg, T. J. Quantum-coherent coupling of a mechanical oscillator to an optical cavity mode. *Nature* **482**, 63–67 (2012).
- Taylor, J. M., Sørensen, A. S., Marcus, C. M. & Polzik, E. S. Laser cooling and optical detection of excitations in a LC electrical circuit. *Phys. Rev. Lett.* **107**, 273601 (2011).
- Regal, C. A. & Lehnert, K. W. From cavity electromechanics to cavity optomechanics. *J. Phys. Conf. Ser.* **264**, 012025 (2011).

10. Safavi-Naeini, A. H. & Painter, O. Proposal for an optomechanical traveling wave phonon-photon translator. *New J. Phys.* **13**, 013017 (2011).
11. Wang, Y.-D. & Clerk, A. A. Using interference for high fidelity quantum state transfer in optomechanics. *Phys. Rev. Lett.* **108**, 153603 (2012).
12. Stannigel, K., Rabl, P., Sørensen, A. S., Zoller, P. & Lukin, M. D. Optomechanical transducers for long-distance quantum communication. *Phys. Rev. Lett.* **105**, 220501 (2010).
13. Dong, C., Fiore, V., Kuzyk, M. C. & Wang, H. Optomechanical dark mode. *Science* **338**, 1609–1613 (2012).
14. Palomaki, T. A., Harlow, J. W., Teufel, J. D., Simmonds, R. W. & Lehnert, K. W. Coherent state transfer between itinerant microwave fields and a mechanical oscillator. *Nature* **495**, 210–214 (2013).
15. Bochmann, J., Vainsencher, A., Awschalom, D. D. & Cleland, A. N. Nanomechanical coupling between microwave and optical photons. *Nature Phys.* **9**, 712–716 (2013).
16. Thompson, J. D. *et al.* Strong dispersive coupling of a high finesse cavity to a micromechanical membrane. *Nature* **452**, 72–75 (2008).
17. Andrews, R. W. *et al.* Reversible and efficient conversion between microwave and optical light. Preprint at <http://arxiv.org/abs/1310.5276> (2013).
18. Truitt, P. A., Hertzberg, J. B., Huang, C. C., Ekinci, K. L. & Schwab, K. C. Efficient and sensitive capacitive readout of nanomechanical resonator arrays. *Nano Lett.* **7**, 120–126 (2007).
19. Unterreithmeier, Q. P., Weig, E. M. & Kotthaus, J. P. Universal transduction scheme for nanomechanical systems based on dielectric forces. *Nature* **458**, 1001–1004 (2009).
20. Weis, S. *et al.* Optomechanically induced transparency. *Science* **330**, 1520–1523 (2010).
21. Ilchenko, V. S. *et al.* K_a -band all-resonant photonic microwave receiver. *IEEE Photonics Technol. Lett.* **20**, 1600–1612 (2008).
22. Devgan, P. S., Pruessner, M. W., Urick, V. J. & Williams, K. J. Detecting low-power and RF signals and using a multimode optoelectronic oscillator and integrated optical filter. *IEEE Photonics Technol. Lett.* **22**, 152–154 (2010).
23. Anetsberger, G. *et al.* Near-field cavity optomechanics with nanomechanical oscillators. *Nature Phys.* **5**, 909–914 (2009).
24. Horowitz, P. & Hill, W. *The Art of Electronics* 2nd edn, 428–454 (Cambridge Univ. Press, 1989).
25. Jensen, K., Weldon, J., Garcia, H. & Zettl, A. Nanotube radio. *Nano Lett.* **7**, 3508–3511 (2007).
26. Massel, F. *et al.* Microwave amplification with nanomechanical resonators. *Nature* **480**, 351–354 (2011).
27. Kraus, J. D. *Radio Astronomy* (McGraw, 1966).
28. Kovacs, H., Moskau, D. & Spraul, M. Cryogenically cooled probes—a leap in NMR technology. *Prog. Nucl. Magn. Reson. Spectrosc.* **46**, 131–155 (2005).
29. Resmer, F., Seton, H. C. & Hutchison, J. M. Cryogenic receive coil and low noise preamplifier for MRI at 0.01 T. *J. Magn. Reson.* **203**, 57–65 (2010).
30. Yu, P.-L., Purdy, T. P. & Regal, C. A. Control of material damping in high-Q membrane microresonators. *Phys. Rev. Lett.* **108**, 083603 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the DARPA project QUASAR, the European Union Seventh Framework Program through SIQS (grant no. 600645) and iQUOEMS (grant no. 323924), and the ERC grants QIOS (grant no. 306576) and INTERFACE (grant no. 291038). We would like to thank J. H. Müller for valuable discussions, A. Barg and A. Næsby for assistance with the interferometer, and L. Jørgensen for cleanroom support.

Author Contributions T.B., A. Simonsen, S.S., K.U. and A. Schliesser performed the experiments and analysed the data. L.G.V. and S.S. designed and fabricated the membrane and the capacitor. J.A. designed the electronic readout circuit. J.M.T., K.U., A. Sørensen, A. Schliesser, E.Z. and E.S.P. developed the model. T.B., A. Schliesser and E.S.P. wrote the paper. A. Schliesser coordinated most of the work. E.S.P. conceived and supervised the project. All authors discussed the results and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.S.P. (polzik@nbi.dk) or A. Schliesser (albert.schliesser@nbi.dk).

Self-assembly of hydrogen-bonded two-dimensional quasicrystals

Natalie A. Wasio¹, Rebecca C. Quardokus¹, Ryan P. Forrest¹, Craig S. Lent², Steven A. Corcelli¹, John A. Christie¹, Kenneth W. Henderson¹ & S. Alex Kandel¹

The process of molecular self-assembly on solid surfaces is essentially one of crystallization in two dimensions, and the structures that result depend on the interplay between intermolecular forces and the interaction between adsorbates and the underlying substrate¹. Because a single hydrogen bond typically has an energy between 15 and 35 kilojoules per mole, hydrogen bonding can be a strong driver of molecular assembly; this is apparent from the dominant role of hydrogen bonding in nucleic-acid base pairing, as well as in the secondary structure of proteins. Carboxylic acid functional groups, which provide two hydrogen bonds, are particularly promising and reliable in creating and maintaining surface order, and self-assembled monolayers of benzoic acids produce structure that depends on the number and relative placement of carboxylic acid groups^{2–6}. Here we use scanning tunnelling microscopy to study self-assembled monolayers of ferrocenecarboxylic acid (FcCOOH), and find that, rather than producing dimeric or linear structures typical of carboxylic acids, FcCOOH forms highly unusual cyclic hydrogen-bonded pentamers, which combine with simultaneously formed FcCOOH dimers to form two-dimensional quasicrystallites that exhibit local five-fold symmetry and maintain translational and rotational order (without periodicity) for distances of more than 400 ångströms.

Quasicrystals—first observed experimentally in 1982—exhibit long-range, non-periodic order, and may contain extended regions with ‘forbidden’ (5, 8, 10 or 12-fold) rotational symmetry^{7,8}. The majority of known quasicrystals are binary or ternary metal alloys such as AlCuMn, or are materials derived from these alloys by adsorption or epitaxy^{9,10}. Apart from metal alloys, other quasicrystalline systems discovered include terpolymer systems¹¹, liquid crystals and micelles^{12,13}, and metal¹⁴ and silica¹⁵ nanoparticle assemblies. FcCOOH, then, represents an entirely new class of quasicrystalline materials, and is the only one in which molecular self-assembly results in five-fold symmetry.

The compound ferroceneacetic acid (FcCH₂COOH) differs from FcCOOH only by the addition of a methylene group between the ferrocene and the carboxylic acid (Fig. 1a) and it shows no evidence of pentamer formation, instead forming monolayers of close-packed dimers. Molecular-resolution images of FcCOOH and FcCH₂COOH monolayers on gold (Au(111)) surfaces acquired using scanning tunnelling microscopy (STM) are presented in Fig. 1b (FcCH₂COOH) and Fig. 1c, d (FcCOOH). The FcCH₂COOH monolayer is ordered and regular, consistent with the formation of doubly hydrogen-bonded dimers, followed by dense packing of these dimers into a lattice. STM resolution of ferrocene groups in covalently bound dimers depends on the relative orientation between the ring system and the surface, with rings parallel to the surface creating much brighter features, and perpendicularly oriented ferrocene much dimmer¹⁶. The presence of both bright and dim features in Fig. 1b shows that FcCH₂COOH, too, has a range of adsorption conformations. Interpretation of the monolayer as consisting of rows of dimers results in an Fe–Fe hydrogen bonding distance of 11.3 Å and an intermolecular spacing of 6.7 Å, consistent with the condensed-phase distances determined crystallographically¹⁷.

The FcCOOH monolayer, however, has no apparent long-range periodicity. Nearly all bright molecular features are grouped into pentagons, though there are some dimers and other defects evident. In addition to areas such as these, where pentamers cover the surface with nearly a complete monolayer, we have also observed areas of the surface with lower adsorbate density; at these lower coverages, isolated pentamers are present (Extended Data Fig. 1a). This result is at odds with the preference for dimer formation observed for monolayers of aliphatic and benzoic carboxylic acids^{3–5}, the structure of the FcCH₂COOH monolayer shown here and the known three-dimensional crystal structure of solid FcCOOH (ref. 17).

Catemers are chains of carboxylic acids in which each molecule hydrogen-bonds to two others; for some acids, catemers constitute either the lowest-energy structure or a metastable structure resulting in lattice polymorphism¹⁸. The degree of hydrogen bonding is the same as for dimer formation, although catemers are generally disfavoured because of non-optimal OH⋯O bond angles, as well as for entropic reasons. We propose that the FcCOOH pentamers observed are cyclic catemers (Fig. 2b). Cyclization would at first seem to create even more bond strain in OH⋯O hydrogen bonds, and impose a higher cost in entropy. However, this is offset by an additional favourable interaction between the carbonyl group and the hydrogen atom on the second position of

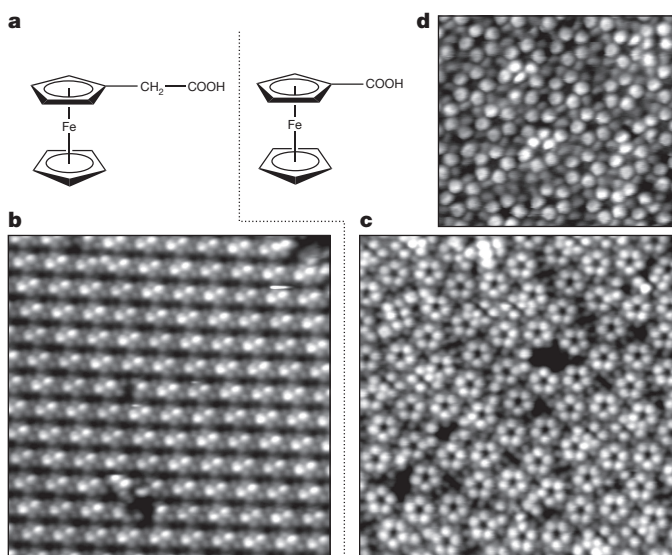


Figure 1 | STM images of FcCOOH and FcCH₂COOH. **a**, Structure of FcCH₂COOH and FcCOOH acid. **b**, **c**, Rows of dimers are observed for FcCH₂COOH (**b**), whereas the structure of FcCOOH (**c**) is dominated by pentamers; both images are 205 Å × 205 Å. **d**, Higher-resolution, 110 Å × 100 Å image showing submolecular structure in the FcCOOH Cp rings; dimer features packed between pentamers also appear more clearly in this image.

¹Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA. ²Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA.

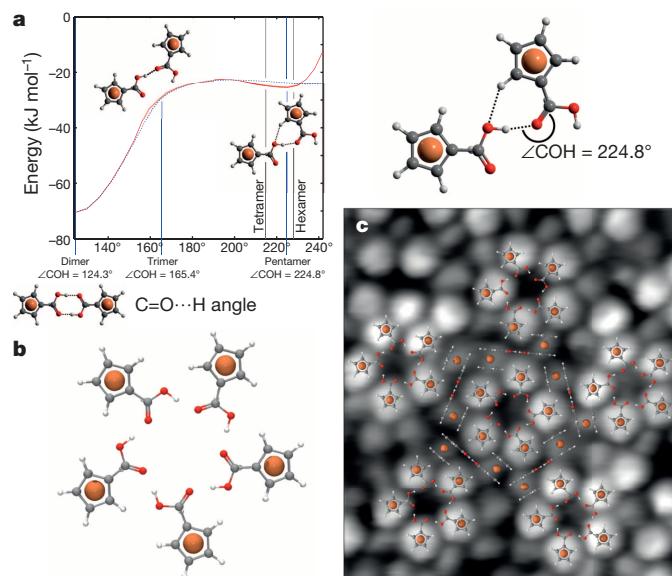


Figure 2 | Calculations and proposed structures for FcCOOH dimers and pentamers. **a**, Calculated energies for FcCOOH dimers as a function of bond angle $\angle\text{COH}$ (red) compared with a similar calculation for HCOOH dimers (blue). Both molecules have the same deep well for the doubly bonded dimer, but only FcCOOH has the additional well at $\angle\text{COH} = 225^\circ$ from $\text{CH}\cdots\text{O}$ hydrogen bonding. Thin vertical lines show the angles necessary for formation of cyclic structures with three to six FcCOOH molecules. **b**, Minimum-energy structure of (FcCOOH)₅ calculated by DFT. **c**, Model structure, with the central pentamer surrounded by five FcCOOH dimers; the pentamer has Cp rings parallel to the surface and appears bright in the images, whereas the perpendicular Cp rings of the dimer result in much dimmer features. The model is compared with the image in Fig. 1d, scaled and rotated, and slightly stretched to correct for distortion in the STM image.

the cyclopentadienyl (Cp) ring. Bringing these atoms into proximity allows for the formation of a $\text{CH}\cdots\text{O}$ hydrogen bond. Such bonds are typically less than 10 kJ mol^{-1} in strength, and although they are not generally considered in small-molecule interactions, they are known to influence lattice ordering in crystals, and are important in the structure and function of large biomolecules¹⁹. Although $\text{CH}\cdots\text{O}$ hydrogen bonds are not evident in the crystal structure of FcCOOH, there is evidence that $\text{CH}\cdots\text{N}$ bonding occurs in the solid-state structure of FcCN (ref. 20).

We performed density functional theory (DFT) calculations to determine whether $\text{CH}\cdots\text{O}$ interactions were sufficient to offset the bending of $\text{OH}\cdots\text{O}$ bonds. In the first set of calculations, the geometry of a FcCOOH dimer with two hydrogen bonds was optimized, and the minimum-energy structure found to have a 124.3° angle between the intermolecular hydrogen bond and the $\text{C}=\text{O}$ bond direction of either carbonyl. The angle $\angle\text{COH}$ was varied to determine its effect on the dimer energy, constraining the dimer to be bent while re-optimizing all other aspects of the geometry. The results are shown in Fig. 2a. The dimer geometry has a deep well resulting from two hydrogen bonds; as the angle increases, one of these bonds is broken and the other strained, with energy increasing until the angle reaches 195° . At this point, the favourable $\text{CH}\cdots\text{O}$ interaction outweighs the cost of further bending the $\text{OH}\cdots\text{O}$ bond, leading to a 2.5 kJ mol^{-1} well at $\angle\text{COH} = 225^\circ$. This closely matches the 224.8° angle necessary to form a cyclic pentamer. Furthermore, the COH bond angles for cyclic tetramers or hexamers are displaced from this energy minimum, which supports the experimentally observed predominance of five-membered rings. We repeated the energy calculations using second-order Møller–Plesset perturbation theory with the DFT-optimized geometry, and found an angular dependence similar to that seen in Fig. 2, including a local minimum at 225° .

The blue dotted line in Fig. 2a shows a similar calculation for formic acid (HCOOH) dimers, with the curve scaled to overlay the FcCOOH dimer curve. The functional forms of these curves are nearly identical between $\angle\text{COH} = 125^\circ$ and 200° , indicating that the underlying processes—the breaking of one hydrogen bond and the bending of the other—are largely independent of the rest of the molecule. The well caused by $\text{CH}\cdots\text{O}$ hydrogen bonding, however, is completely absent for formic acid, because the lone hydrogen on the formic acid carbon is never close enough to an oxygen to participate in any intermolecular interaction.

The $\text{CH}\cdots\text{O}$ well at $\angle\text{COH} = 225^\circ$ in Fig. 2a appears shallow, but this is in part because it is a calculation for a strained dimer rather than for a full cyclic pentamer. The strained dimer makes one $\text{OH}\cdots\text{O}$ hydrogen bond and one $\text{CH}\cdots\text{O}$ bond, but leaves unbonded the $\text{C}=\text{O}$ on one molecule and the COH on the other. In the cyclic pentamer, all $\text{C}=\text{O}$ and COH groups participate in hydrogen bonding, potentially doubling the cohesive energy per molecule relative to what is shown in Fig. 2a. To determine more accurately the relative stability of the dimer versus the cyclic catemers, we performed additional DFT calculations to obtain energies and optimized geometries for FcCOOH dimers and for hydrogen-bonded rings containing three to six molecules (Extended Data Table 1). In all cases, the pentamer was found to be more stable than any other cyclic structure; of these, the hexamer was the next most stable. The relative stability of the dimer and pentamer depends on the choice of functionals used for the calculation, with B3LYP and PBE predicting the dimer to be more stable and M06-L and M06-2X favouring the pentamer. The two structures are close enough in energy—less than 2 kJ mol^{-1} for all functionals except M06-2X—that both should be observable given that $k_{\text{B}}T$ (k_{B} , Boltzmann's constant; T , temperature) ranges from 0.64 to 2.49 kJ mol^{-1} , depending on the point between preparation (298 K) and imaging (77 K) when the structure becomes kinetically frozen.

For FcCH_2COOH molecules, the length of the additional methylene group prevents any interaction between the carboxyl oxygen and the Cp hydrogens. If other factors—for example steric packing of the ferrocene Cp rings—were responsible for pentamer formation, both FcCOOH and FcCH_2COOH would aggregate into pentamers in a similar fashion. In contrast, the observation of dimers and a complete absence of pentamers for FcCH_2COOH provides strong corroborating evidence that it is the $\text{CH}\cdots\text{O}$ bonds that stabilize FcCOOH pentamers. Other explanations, such as registry to the underlying surface, also do not explain the observations: the Au(111) substrate has hexagonal symmetry, and to the extent that benzene remains on the surface, it packs in a rectangular lattice under our experimental conditions²¹. Furthermore, the isolated pentamers observed in Extended Data Fig. 1a are randomly oriented in two dimensions, which rules out the corrugation of the substrate as a driving cause of pentamer formation.

In addition to pentamers, the FcCOOH monolayer contains a number of other bright features that can be resolved in Fig. 1c, d, including dimers and groups of dimers as well as individual bright features. There is also a large number of less bright features between pentamers, which are very dim in Fig. 1c and clearer in Fig. 1d. As for the dim FcCH_2COOH molecules in Fig. 1b, we interpret these as FcCOOH molecules with Cp rings tilted with respect to the surface. A possible structural model is shown in Fig. 2c: the central pentamer is surrounded by five dimers, which well match the observed dim features. This arrangement creates multiple $\text{CH}\cdots\pi$ interactions, which are known to drive ordering in similar systems²². The dimers thus act as ‘glue’ between adjacent pentamers—whereas the brighter pentamer features seem to align vertex to vertex, the dimers create an effectively larger pentagonal unit that then packs edge to edge with its neighbours.

We note that the molecular arrangement in Fig. 1c seems to have a single chirality; the effect is subtle but consistent, and the chirality reverses for the area imaged in Fig. 1d. For neighbouring pentamers aligned vertex to vertex, the two pentagons are offset so that the pair has point symmetry but not plane symmetry; that is, the pentagons’ symmetry

axes may be parallel but not collinear. This observation is explained in the model, because the *trans*-configuration of the dimers in the model in Fig. 2c create handedness that will propagate to create long-range chirality in the monolayer. In some cases, monolayer defects create grain boundaries that interrupt the long-range structure. Both orientation and chirality can change across these boundaries; an example is shown in Extended Data Fig. 5, along with diagrams to illustrate the chiral structure.

The packing of pentamers to form a FcCOOH monolayer results in quasicrystalline structure, which is analysed in Fig. 3. In Fig. 3a, a $405 \text{ \AA} \times 360 \text{ \AA}$ image of a FcCOOH monolayer is low-pass-filtered to de-emphasize the internal structure of the pentamers and thus emphasize the longer-range order (the raw data are presented in Extended Data Fig. 5d). Each pentamer is then overlaid with a pentagon that—relative to the pentagon formed with molecules at vertices—was expanded in size and rotated by 180° . The image is patterned but has no apparent periodicity, and the pentagons and their interstitial spaces match the shapes in a Penrose 'P1' tiling²³ (pentagon, rhombus, boat and star), with only a few small gaps in evidence. A number of intermediate-range structures exactly mirror those of a Penrose tiling, which is drawn in Fig. 3b for comparison. As in the Penrose tiling, orientational ordering of pentagons in the experimental image is nearly perfect, with only two directions observed; this is also presented quantitatively in Extended Data Figs 1b and 2.

Local five-fold and ten-fold symmetry of the FcCOOH monolayer can be seen in the two-dimensional Fourier transform in Fig. 3c and the two-dimensional spatial correlation function in Fig. 3d; the data set used to produce both of these is in Extended Data Fig. 1b. The Fourier transform (Extended Data Fig. 3a) has ten-fold-symmetric peaks at a number of radii, and both the sharpness and the symmetry are characteristic of quasicrystalline ordering. The spatial correlation function (Extended Data Fig. 3b) maps displacement vectors between every pair of pentagons in an image; it, too, shows ten-fold symmetry as well as long-range translational order. This order is quasiperiodic, however, because it corresponds to a sum of multiple frequencies related by irrational ratios: in this case, powers of the golden ratio, $\phi = (1 + \sqrt{5})/2$ (Extended Data Fig. 4). Quasiperiodicity and orientational order can be considered the essential properties of quasicrystals²⁴. FcCOOH

monolayers are also quasicrystalline by an alternative definition, which requires an essentially sharp diffraction pattern and the absence of translational periodicity²⁵. The relatively small size of ordered regions ($>400 \text{ \AA}$) is fairly typical for self-assembled monolayers with hexagonal or other allowed symmetry. FcCOOH might be more accurately described, therefore, as polyquasicrystalline, and the regions imaged here, as quasicrystallites; this is largely a matter of terminology, however.

Differences between the Penrose tiling and the FcCOOH monolayer include the small gaps described above and the presence of relatively more rhombuses and fewer stars, as well as the prevalence of pentagons with four other pentagons immediately adjacent. In a Penrose tiling, pentagons may have two, three or five immediate neighbours, and two adjacent pentagons cannot have the same number of neighbours. Replicating these constraints would be difficult with pairwise interactions, and this probably explains differences between the monolayer and Penrose structures. However, the intermediate-range order of the monolayer matches that of the Penrose tiling quite well.

Maximal packing cannot be the sole driving force in the formation of a quasicrystalline monolayer. Higher-density packing of pentagons is possible with periodic lattices; indeed, other investigations of the self-assembly of molecules with C_5 symmetry uniformly have found monolayers with periodic lattice structures^{26,27}, including a complex (but still periodic) Archimedean tiling²⁸. Random packing of pentagons with the constraint that edges and vertices must match, however, produces patterns that appear much more like quasicrystals²⁹. The structures in ref. 29 are more dendritic and less dense than those we observe for FcCOOH. Physically, we would expect thermal motions during deposition to allow for some reorientation, increasing the packing density until the overall structure either reaches a free-energy minimum or becomes kinetically locked. The result is similar to that predicted when a Metropolis algorithm is used to anneal structure after random growth³⁰.

The spontaneous pentamer formation observed for FcCOOH will probably also occur for molecules functionalized at the 3 and 4 positions—that is, not immediately adjacent to the COOH group; some functionalizations of the other Cp ring may also not hinder pentamer formation. There is potentially a large range of novel supramolecular assemblies based on these pentameric building blocks, in which the chemistry of side groups can control the interaction of pentamers and influence the formation of crystalline or quasicrystalline structure.

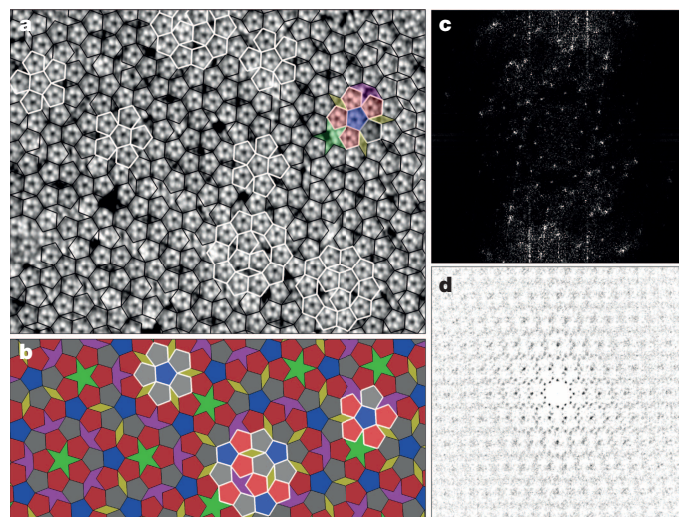


Figure 3 | Comparison of a FcCOOH monolayer structure to a P1 Penrose tiling. **a**, $405 \text{ \AA} \times 360 \text{ \AA}$ image, overlaid with pentagons to show the long-range, quasicrystalline order. **b**, Groups of pentagons highlighted in white in **a** show some structural motifs common to both the experimental data and a Penrose P1 tiling; the interstitial spaces between pentamers in **a** also match the star, boat and rhombus of the P1 tiling. **c**, Two-dimensional Fourier transform; the innermost ring arises from long-range five-fold and ten-fold symmetry. **d**, This symmetry can also be seen in the two-dimensional spatial correlation function.

METHODS SUMMARY

Solutions of approximately 10 mM of FcCOOH and FcCH_2COOH acid were prepared in an Ar-purged glovebox, using benzene as the solvent. Small quantities of the ferrocene derivatives were loaded into a pulsed solenoid valve and deposited in vacuum onto an Ar-sputtered and annealed Au(111)-on-mica sample at room temperature ($\sim 295 \text{ K}$). Once deposited, the sample was immediately transferred into an ultrahigh-vacuum chamber (10^{-10} Torr) at 77 K, and STM images were acquired under constant-current conditions. Image filtering (high pass in the scan direction, or low pass overall) was done in some cases to decrease noise, but no other frequency-selective filter was used. DFT calculations were performed using the Q-CHEM software package³¹. The 6-311++G** basis set was used for all C, H and O atoms, and the LANL2TZ effective core potential basis set was used for all Fe atoms.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 August; accepted 12 December 2013.

- Smith, R., Lewis, P. & Weiss, P. Patterning self-assembled monolayers. *Prog. Surf. Sci.* **75**, 1–68 (2004).
- Lackinger, M. & Heckl, W. Carboxylic acids: versatile building blocks and mediators for two-dimensional supramolecular self-assembly. *Langmuir* **25**, 11307–11321 (2009).
- Clair, S. *et al.* STM study of terephthalic acid self-assembly on Au(111): hydrogen-bonded sheets on an inhomogeneous substrate. *J. Phys. Chem. B* **108**, 14585–14590 (2004).
- Fuhr, J. D. *et al.* Interplay between hydrogen bonding and molecule–substrate interactions in the case of terephthalic acid molecules on Cu(001) surfaces. *J. Phys. Chem. C* **117**, 1287–1296 (2013).

5. Heininger, C., Kampschulte, L., Heckl, W. & Lackinger, M. Distinct differences in self-assembly of aromatic linear dicarboxylic acids. *Langmuir* **25**, 968–972 (2009).
6. Ye, Y. *et al.* A unified model: self-assembly of trimesic acid on gold. *J. Phys. Chem. C* **111**, 10138–10141 (2007).
7. Levine, D. & Steinhardt, P. Quasicrystals—a new class of ordered structures. *Phys. Rev. Lett.* **53**, 2477–2480 (1984).
8. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.* **53**, 1951–1953 (1984).
9. Smerdon, J. A. *et al.* Nucleation and growth of a quasicrystalline monolayer: bi-adsorption on the fivefold surface of $i\text{-Al}(70)\text{Pd}(21)\text{Mn}(9)$. *Phys. Rev. B* **78**, 075407 (2008).
10. Gierer, M. *et al.* Fivefold surface of quasicrystalline AlPdMn : structure determination using low-energy-electron diffraction. *Phys. Rev. B* **57**, 7628–7641 (1998).
11. Hayashida, K., Dotera, T., Takano, A. & Matsushita, Y. Polymeric quasicrystal: mesoscopic quasicrystalline tiling in ABC star polymers. *Phys. Rev. Lett.* **98**, 195502 (2007).
12. Fischer, S. *et al.* Colloidal quasicrystals with 12-fold and 18-fold diffraction symmetry. *Proc. Natl Acad. Sci. USA* **108**, 1810–1814 (2011).
13. Zeng, X. *et al.* Supramolecular dendritic liquid quasicrystals. *Nature* **428**, 157–160 (2004).
14. Talapin, D. V. *et al.* Quasicrystalline order in self-assembled binary nanoparticle superlattices. *Nature* **461**, 964–967 (2009).
15. Xiao, C., Fujita, N., Miyasaka, K., Sakamoto, Y. & Terasaki, O. Dodecagonal tiling in mesoporous silica. *Nature* **487**, 349–353 (2012).
16. Zhong, D. *et al.* Multilevel supramolecular architectures self-assembled on metal surfaces. *ACS Nano* **4**, 1997–2002 (2010).
17. Cotton, F. & Reid, A. Solid-state structure of ferrocenecarboxylic acid, $[\text{Fe}(\text{C}_5\text{H}_4\text{CO}_2\text{H})(\text{C}_5\text{H}_5)]$. *Acta Crystallogr.* **41**, 686–688 (1985).
18. Beyer, T. & Price, S. Dimer or catemer? Low-energy crystal packings for small carboxylic acids. *J. Phys. Chem. B* **104**, 2647–2655 (2000).
19. Steiner, T. & Desiraju, G. Distinction between the weak hydrogen bond and the van der Waals interaction. *Chem. Commun.* 891–892 (1998).
20. Bell, W., Ferguson, G. & Glidewell, C. Cyanoferrocene: a two-dimensional network generated by short $\text{C-H}\cdots\text{N}$ hydrogen bonds. *Acta Crystallogr. C* **52**, 1928–1930 (1996).
21. Quardokus, R. *et al.* Adsorption of diferrocenylacetylene on $\text{Au}(111)$ studied by scanning tunneling microscopy. *Phys. Chem. Chem. Phys.* **15**, 6973–6981 (2013).
22. Nishio, M. The CH/π hydrogen bond in chemistry. Conformation, supramolecules, optical resolution and interactions involving carbohydrates. *Phys. Chem. Chem. Phys.* **13**, 13873–13900 (2011).
23. Penrose, R. Pentaplexity: a class of non-periodic tilings of the plane. *Eureka* **39**, 16–32 (1978).
24. Levine, D. & Steinhardt, P. Quasi-crystals: 1. Definition and structure. *Phys. Rev. B* **34**, 596–616 (1986).
25. International Union of Crystallography. Report of the Executive Committee for 1991. *Acta Crystallogr. A* **6**, 922–946 (1992).
26. Bauert, T. *et al.* Building 2D crystals from 5-fold-symmetric molecules. *J. Am. Chem. Soc.* **131**, 3460–3461 (2009).
27. Guillermet, O. *et al.* Self-assembly of fivefold-symmetric molecules on a threefold-symmetric surface. *Angew. Chem. Int. Ed.* **48**, 1970–1973 (2009).
28. Eciya, D. *et al.* Five-vertex Archimedean surface tessellation by lanthanide-directed molecular self-assembly. *Proc. Natl Acad. Sci. USA* **110**, 6678–6681 (2013).
29. Bauer, K. & Krey, U. Model-calculations on random pentagon clusters compared to 2D-quasicrystals. *Z. Phys. B* **75**, 385–392 (1989).
30. Reichert, M. & Gähler, F. Cluster model of decagonal tilings. *Phys. Rev. B* **68**, 214202 (2003).
31. Shao, Y. *et al.* Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **8**, 3172–3191 (2006).

Acknowledgements This work was supported by the US National Science Foundation under grant NSF CHE-1124762. The authors acknowledge high-performance computing resources and support from the Center for Research Computing at the University of Notre Dame.

Author Contributions N.A.W. and R.C.Q. performed STM measurements and made the initial experimental discovery. R.P.F., C.S.L. and S.A.C. designed the computational approach, and R.P.F. performed the calculations. J.A.C. and K.W.H. provided the molecules used, as well as several others used as tests and controls. All authors participated in analysis and interpretation of the results. N.A.W. drafted the manuscript. S.A.K. prepared extended data figures, edited the manuscript and coordinated the efforts of the research team.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.K. (skandel@nd.edu).

Rate of tree carbon accumulation increases continuously with tree size

N. L. Stephenson¹, A. J. Das¹, R. Condit², S. E. Russo³, P. J. Baker⁴, N. G. Beckman^{3†}, D. A. Coomes⁵, E. R. Lines⁶, W. K. Morris⁷, N. Rüger^{2,8†}, E. Álvarez⁹, C. Blundo¹⁰, S. Bunyavejchewin¹¹, G. Chuyong¹², S. J. Davies¹³, Á. Duque¹⁴, C. N. Ewango¹⁵, O. Flores¹⁶, J. F. Franklin¹⁷, H. R. Grau¹⁰, Z. Hao¹⁸, M. E. Harmon¹⁹, S. P. Hubbell^{2,20}, D. Kenfack¹³, Y. Lin²¹, J.-R. Makana¹⁵, A. Malizia¹⁰, L. R. Malizia²², R. J. Pabst¹⁹, N. Pongpattananurak²³, S.-H. Su²⁴, I.-F. Sun²⁵, S. Tan²⁶, D. Thomas²⁷, P. J. van Mantgem²⁸, X. Wang¹⁸, S. K. Wiser²⁹ & M. A. Zavala³⁰

Forests are major components of the global carbon cycle, providing substantial feedback to atmospheric greenhouse gas concentrations¹. Our ability to understand and predict changes in the forest carbon cycle—particularly net primary productivity and carbon storage—increasingly relies on models that represent biological processes across several scales of biological organization, from tree leaves to forest stands^{2,3}. Yet, despite advances in our understanding of productivity at the scales of leaves and stands, no consensus exists about the nature of productivity at the scale of the individual tree^{4–7}, in part because we lack a broad empirical assessment of whether rates of absolute tree mass growth (and thus carbon accumulation) decrease, remain constant, or increase as trees increase in size and age. Here we present a global analysis of 403 tropical and temperate tree species, showing that for most species mass growth rate increases continuously with tree size. Thus, large, old trees do not act simply as senescent carbon reservoirs but actively fix large amounts of carbon compared to smaller trees; at the extreme, a single big tree can add the same amount of carbon to the forest within a year as is contained in an entire mid-sized tree. The apparent paradoxes of individual tree growth increasing with tree size despite declining leaf-level^{8–10} and stand-level¹⁰ productivity can be explained, respectively, by increases in a tree's total leaf area that outpace declines in productivity per unit of leaf area and, among other factors, age-related reductions in population density. Our results resolve conflicting assumptions about the nature of tree growth, inform efforts to understand and model forest carbon dynamics, and have additional implications for theories of resource allocation¹¹ and plant senescence¹².

A widely held assumption is that after an initial period of increasing growth, the mass growth rate of individual trees declines with increasing tree size^{4,5,13–16}. Although the results of a few single-species studies have been consistent with this assumption¹⁵, the bulk of evidence cited in support of declining growth is not based on measurements of individual tree mass growth. Instead, much of the cited evidence documents either the well-known age-related decline in net primary productivity (hereafter 'productivity') of even-aged forest stands¹⁰ (in which the trees are all of a similar age) or size-related declines in the rate of mass gain per

unit leaf area (or unit leaf mass)^{8–10}, with the implicit assumption that declines at these scales must also apply at the scale of the individual tree. Declining tree growth is also sometimes inferred from life-history theory to be a necessary corollary of increasing resource allocation to reproduction^{11,16}. On the other hand, metabolic scaling theory predicts that mass growth rate should increase continuously with tree size⁶, and this prediction has also received empirical support from a few site-specific studies^{6,7}. Thus, we are confronted with two conflicting generalizations about the fundamental nature of tree growth, but lack a global assessment that would allow us to distinguish clearly between them.

To fill this gap, we conducted a global analysis in which we directly estimated mass growth rates from repeated measurements of 673,046 trees belonging to 403 tropical, subtropical and temperate tree species, spanning every forested continent. Tree growth rate was modelled as a function of log(tree mass) using piecewise regression, where the independent variable was divided into one to four bins. Conjoined line segments were fitted across the bins (Fig. 1).

For all continents, aboveground tree mass growth rates (and, hence, rates of carbon gain) for most species increased continuously with tree mass (size) (Fig. 2). The rate of mass gain increased with tree mass in each model bin for 87% of species, and increased in the bin that included the largest trees for 97% of species; the majority of increases were statistically significant (Table 1, Extended Data Fig. 1 and Supplementary Table 1). Even when we restricted our analysis to species achieving the largest sizes (maximum trunk diameter >100 cm; 33% of species), 94% had increasing mass growth rates in the bin that included the largest trees. We found no clear taxonomic or geographic patterns among the 3% of species with declining growth rates in their largest trees, although the small number of these species (thirteen) hampers inference. Declining species included both angiosperms and gymnosperms in seven of the 76 families in our study; most of the seven families had only one or two declining species and no family was dominated by declining species (Supplementary Table 1).

When we log-transformed mass growth rate in addition to tree mass, the resulting model fits were generally linear, as predicted by metabolic scaling theory⁶ (Extended Data Fig. 2). Similar to the results of our main

¹US Geological Survey, Western Ecological Research Center, Three Rivers, California 93271, USA. ²Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Republic of Panama. ³School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588, USA. ⁴Department of Forest and Ecosystem Science, University of Melbourne, Victoria 3121, Australia. ⁵Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. ⁶Department of Geography, University College London, London WC1E 6BT, UK. ⁷School of Botany, University of Melbourne, Victoria 3010, Australia. ⁸Spezielle Botanik und Funktionelle Biodiversität, Universität Leipzig, 04103 Leipzig, Germany. ⁹Jardín Botánico de Medellín, Calle 73, No. 51D-14, Medellín, Colombia. ¹⁰Instituto de Ecología Regional, Universidad Nacional de Tucumán, 4107 Yerba Buena, Tucumán, Argentina. ¹¹Research Office, Department of National Parks, Wildlife and Plant Conservation, Bangkok 10900, Thailand. ¹²Department of Botany and Plant Physiology, Buea, Southwest Province, Cameroon. ¹³Smithsonian Institution Global Earth Observatory—Center for Tropical Forest Science, Smithsonian Institution, PO Box 37012, Washington, DC 20013, USA. ¹⁴Universidad Nacional de Colombia, Departamento de Ciencias Forestales, Medellín, Colombia. ¹⁵Wildlife Conservation Society, Kinshasa/Gombe, Democratic Republic of the Congo. ¹⁶Unité Mixte de Recherche—Peuplements Végétaux et Bioagresseurs en Milieu Tropical, Université de la Réunion/CIRAD, 97410 Saint Pierre, France. ¹⁷School of Environmental and Forest Sciences, University of Washington, Seattle, Washington 98195, USA. ¹⁸State Key Laboratory of Forest and Soil Ecology, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110164, China. ¹⁹Department of Forest Ecosystems and Society, Oregon State University, Corvallis, Oregon 97331, USA. ²⁰Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095, USA. ²¹Department of Life Science, Tunghai University, Taichung City 40704, Taiwan. ²²Facultad de Ciencias Agrarias, Universidad Nacional de Jujuy, 4600 San Salvador de Jujuy, Argentina. ²³Faculty of Forestry, Kasetsart University, Chatuchak Bangkok 10900, Thailand. ²⁴Taiwan Forestry Research Institute, Taipei 10066, Taiwan. ²⁵Department of Natural Resources and Environmental Studies, National Dong Hwa University, Hualien 97401, Taiwan. ²⁶Sarawak Forestry Department, Kuching, Sarawak 93660, Malaysia. ²⁷Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331, USA. ²⁸US Geological Survey, Western Ecological Research Center, Arcata, California 95521, USA. ²⁹Landcare Research, PO Box 40, Lincoln 7640, New Zealand. ³⁰Forest Ecology and Restoration Group, Department of Life Sciences, University of Alcalá, Alcalá de Henares, 28805 Madrid, Spain. †Present addresses: Mathematical Biosciences Institute, Ohio State University, Columbus, Ohio 43210, USA (N.G.B.); German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, 04103 Leipzig, Germany (N.R.).

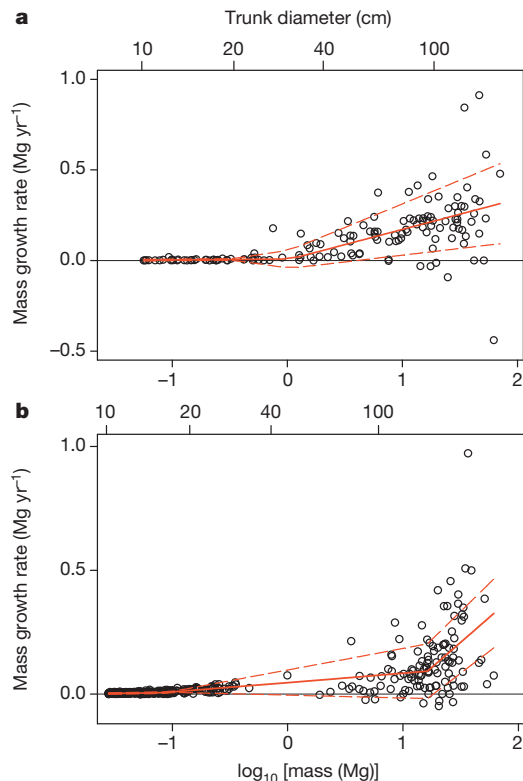


Figure 1 | Example model fits for tree mass growth rates. The species shown are the angiosperm species (*Lecomtedoxa klaineana*, Cameroon, 142 trees) (a) and gymnosperm species (*Picea sitchensis*, USA, 409 trees) (b) in our data set that had the most massive trees (defined as those with the greatest cumulative aboveground dry mass in their five most massive trees). Each point represents a single tree; the solid red lines represent best fits selected by our model; and the dashed red lines indicate one standard deviation around the predicted values.

analysis using untransformed growth, of the 381 log-transformed species analysed (see Methods), the log-transformed growth rate increased in the bin containing the largest trees for 96% of species.

In absolute terms, trees 100 cm in trunk diameter typically add from 10 kg to 200 kg of aboveground dry mass each year (depending on species), averaging 103 kg per year. This is nearly three times the rate for trees of the same species at 50 cm in diameter, and is the mass equivalent to adding an entirely new tree of 10–20 cm in diameter to the forest each year. Our findings further indicate that the extraordinary growth recently reported in an intensive study of large *Eucalyptus regnans* and *Sequoia sempervirens*⁷, which included some of the world's most massive individual trees, is not a phenomenon limited to a few unusual species. Rather, rapid growth in giant trees is the global norm, and can exceed 600 kg per year in the largest individuals (Fig. 3).

Our data set included many natural and unmanaged forests in which the growth of smaller trees was probably reduced by asymmetric competition with larger trees. To explore the effects of competition, we calculated mass growth rates for 41 North American and European species that had published equations for diameter growth rate in the absence of competition. We found that, even in the absence of competition, 85% of the species had mass growth rates that increased continuously with tree size (Extended Data Fig. 3), with growth curves closely resembling those in Fig. 2. Thus, our finding of increasing growth not only has broad generality across species, continents and forest biomes (tropical, subtropical and temperate), it appears to hold regardless of competitive environment.

Importantly, our finding of continuously increasing growth is compatible with the two classes of observations most often cited as evidence of declining, rather than increasing, individual tree growth: with increasing tree size and age, productivity usually declines at the scales of both tree organs (leaves) and tree populations (even-aged forest stands).

First, although growth efficiency (tree mass growth per unit leaf area or leaf mass) often declines with increasing tree size^{8–10}, empirical observations and metabolic scaling theory both indicate that, on average, total tree leaf mass increases as the square of trunk diameter^{17,18}. A typical tree that experiences a tenfold increase in diameter will therefore undergo a roughly 100-fold increase in total leaf mass and a 50–100-fold

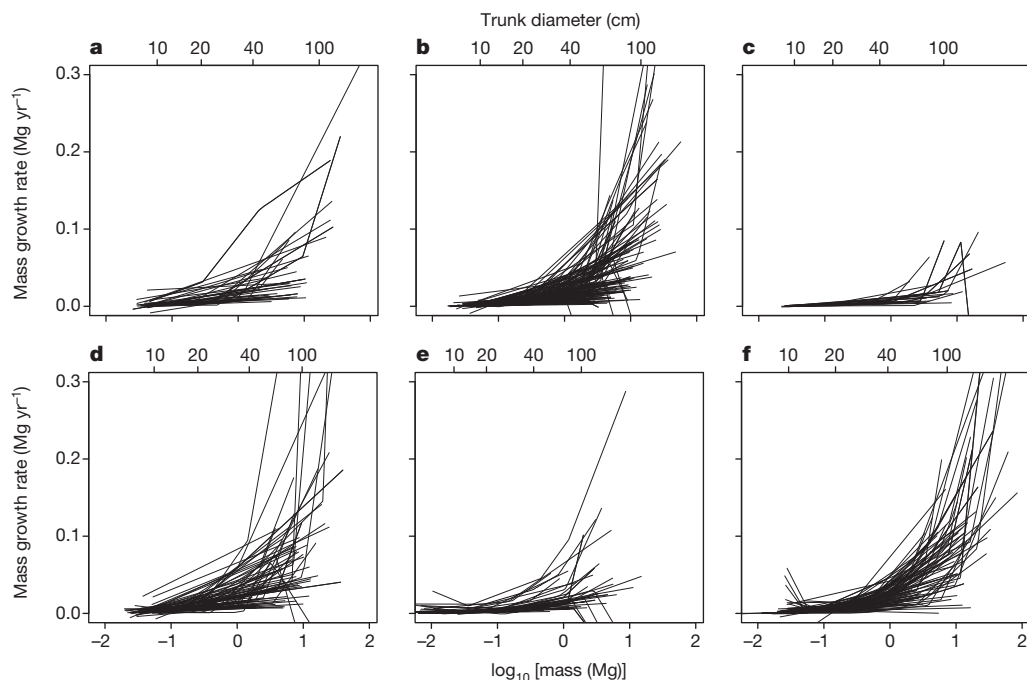


Figure 2 | Aboveground mass growth rates for the 403 tree species, by continent. a, Africa (Cameroon, Democratic Republic of the Congo); b, Asia (China, Malaysia, Taiwan, Thailand); c, Australasia (New Zealand); d, Central and South America (Argentina, Colombia, Panama); e, Europe (Spain); and

f, North America (USA). Numbers of trees, numbers of species and percentages with increasing growth are given in Table 1. Trunk diameters are approximate values for reference, based on the average diameters of trees of a given mass.

Table 1 | Sample sizes and tree growth trends by continent

| Continent | Number of trees | Number of species | Percentage of species with increasing mass growth rate in the largest trees (percentage significant at $P \leq 0.05$) |
|---------------------------|-----------------|-------------------|---|
| Africa | 15,366 | 37 | 100.0 (86.5) |
| Asia | 43,690 | 136 | 96.3 (89.0) |
| Australasia | 45,418 | 22 | 95.5 (95.5) |
| Central and South America | 18,530 | 77 | 97.4 (92.2) |
| Europe | 439,889 | 42 | 90.5 (78.6) |
| North America | 110,153 | 89 | 98.9 (94.4) |
| Total | 673,046 | 403 | 96.8 (89.8) |

The largest trees are those in the last bin fitted by the model. Countries are listed in the legend for Fig. 2.

increase in total leaf area (depending on size-related increases in leaf mass per unit leaf area^{19,20}). Parallel changes in growth efficiency can range from a modest increase (such as in stands where small trees are suppressed by large trees)²¹ to as much as a tenfold decline²², with most changes falling in between^{8,9,19,22}. At one extreme, the net effect of a low (50-fold) increase in leaf area combined with a large (tenfold) decline in growth efficiency would still yield a fivefold increase in individual tree mass growth rate; the opposite extreme would yield roughly a 100-fold increase. Our calculated 52-fold greater average mass growth rate of trees 100 cm in diameter compared to those 10 cm in diameter falls within this range. Thus, although growth efficiency often declines with increasing tree size, increases in a tree's total leaf area are sufficient to overcome this decline and cause whole-tree carbon accumulation rate to increase.

Second, our findings are similarly compatible with the well-known age-related decline in productivity at the scale of even-aged forest stands. Although a review of mechanisms is beyond the scope of this paper^{10,23}, several factors (including the interplay of changing growth efficiency and tree dominance hierarchies²⁴) can contribute to declining productivity at the stand scale. We highlight the fact that increasing individual tree growth rate does not automatically result in increasing stand productivity because tree mortality can drive orders-of-magnitude reductions in population density^{25,26}. That is, even though the large trees in older, even-aged stands may be growing more rapidly, such stands have fewer trees. Tree population dynamics, especially mortality, can thus be a significant contributor to declining productivity at the scale of the forest stand²³.

For a large majority of species, our findings support metabolic scaling theory's qualitative prediction of continuously increasing growth

at the scale of individual trees⁶, with several implications. For example, life-history theory often assumes that tradeoffs between plant growth and reproduction are substantial¹¹. Contrary to some expectations^{11,16}, our results indicate that for most tree species size-related changes in reproductive allocation are insufficient to drive long-term declines in growth rates⁶. Additionally, declining growth is sometimes considered to be a defining feature of plant senescence¹². Our findings are thus relevant to understanding the nature and prevalence of senescence in the life history of perennial plants²⁷.

Finally, our results are relevant to understanding and predicting forest feedbacks to the terrestrial carbon cycle and global climate system¹⁻³. These feedbacks will be influenced by the effects of climatic, land-use and other environmental changes on the size-specific growth rates and size structure of tree populations—effects that are already being observed in forests^{28,29}. The rapid growth of large trees indicates that, relative to their numbers, they could play a disproportionately important role in these feedbacks³⁰. For example, in our western USA old-growth forest plots, trees >100 cm in diameter comprised 6% of trees, yet contributed 33% of the annual forest mass growth. Mechanistic models of the forest carbon cycle will depend on accurate representation of productivity across several scales of biological organization, including calibration and validation against continuously increasing carbon accumulation rates at the scale of individual trees.

METHODS SUMMARY

We estimated aboveground dry mass growth rates from consecutive diameter measurements of tree trunks—typically measured every five to ten years—from long-term monitoring plots. Analyses were restricted to trees with trunk diameter ≥ 10 cm, and to species having ≥ 40 trees in total and ≥ 15 trees with trunk diameter ≥ 30 cm. Maximum trunk diameters ranged from 38 cm to 270 cm among species, averaging 92 cm. We converted each diameter measurement (plus an accompanying height measurement for 16% of species) to aboveground dry mass, M , using published allometric equations. We estimated tree growth rate as $G = \Delta M / \Delta t$ and modelled G as a function of $\log(M)$ for each species using piecewise regression. The independent variable $\log(M)$ was divided into bins and a separate line segment was fitted to G versus $\log(M)$ in each bin so that the line segments met at the bin divisions. Bin divisions were not assigned a priori, but were fitted by the model separately for each species. We fitted models with 1, 2, 3 and 4 bins, and selected the model receiving the most support by Akaike's Information Criterion for each species. Our approach thus makes no assumptions about the shape of the relationship between G and $\log(M)$, and can accommodate increasing, decreasing or hump-shaped relationships. Parameters were fitted with a Gibbs sampler based on Metropolis updates, producing credible intervals for model parameters and growth rates at any diameter; uninformative priors were used for all parameters. We tested extensively for bias, and found no evidence that our results were influenced by model fits failing to detect a final growth decline in the largest trees, possible biases introduced by the 47% of species for which we combined data from several plots, or possible biases introduced by allometric equations (Extended Data Figs 4 and 5).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 August; accepted 27 November 2013.

Published online 15 January 2014.

1. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).

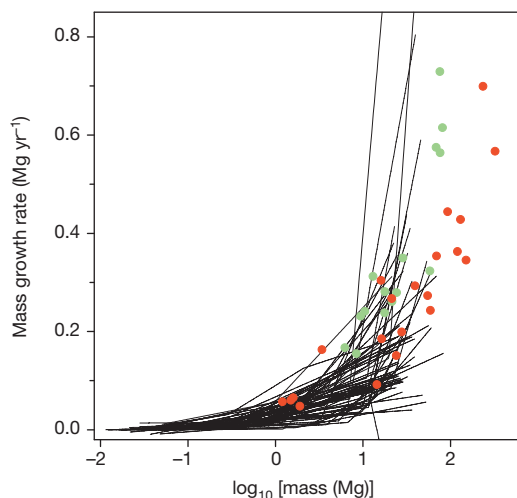


Figure 3 | Aboveground mass growth rates of species in our data set compared with *E. regnans* and *S. sempervirens*. For clarity, only the 58 species in our data set having at least one tree exceeding 20 Mg are shown (lines). Data for *E. regnans* (green dots, 15 trees) and *S. sempervirens* (red dots, 21 trees) are from an intensive study that included some of the most massive individual trees on Earth⁷. Both axes are expanded relative to those of Fig. 2.

2. Medvigy, D., Wofsy, S. C., Munger, J. W., Hollinger, D. Y. & Moorcroft, P. R. Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. *J. Geophys. Res.* **114**, G01002 (2009).
3. Caspersen, J. P., Vanderwel, M. C., Cole, W. G. & Purves, D. W. How stand productivity results from size- and competition-dependent growth and mortality. *PLoS ONE* **6**, e28660 (2011).
4. Kutsch, W. L. *et al.* in *Old-Growth Forests: Function, Fate and Value* (eds Wirth, C., Gleixner, G. & Heimann, M.) 57–79 (Springer, 2009).
5. Meinzer, F. C., Lachenbruch, B. & Dawson, T. E. (eds) *Size- and Age-Related Changes in Tree Structure and Function* (Springer, 2011).
6. Enquist, B. J., West, G. B., Charnov, E. L. & Brown, J. H. Allometric scaling of production and life-history variation in vascular plants. *Nature* **401**, 907–911 (1999).
7. Sillett, S. C. *et al.* Increasing wood production through old age in tall trees. *For. Ecol. Manage.* **259**, 976–994 (2010).
8. Mencuccini, M. *et al.* Size-mediated ageing reduces vigour in trees. *Ecol. Lett.* **8**, 1183–1190 (2005).
9. Drake, J. E., Raetz, L. M., Davis, S. C. & DeLucia, E. H. Hydraulic limitation not declining nitrogen availability causes the age-related photosynthetic decline in loblolly pine (*Pinus taeda* L.). *Plant Cell Environ.* **33**, 1756–1766 (2010).
10. Ryan, M. G., Binkley, D. & Fownes, J. H. Age-related decline in forest productivity: pattern and process. *Adv. Ecol. Res.* **27**, 213–262 (1997).
11. Thomas, S. C. in *Size- and Age-Related Changes in Tree Structure and Function* (eds Meinzer, F. C., Lachenbruch, B. & Dawson, T. E.) 33–64 (Springer, 2011).
12. Thomas, H. Senescence, ageing and death of the whole plant. *New Phytol.* **197**, 696–711 (2013).
13. Carey, E. V., Sala, A., Keane, R. & Callaway, R. M. Are old forests underestimated as global carbon sinks? *Glob. Change Biol.* **7**, 339–344 (2001).
14. Phillips, N. G., Buckley, T. N. & Tissue, D. T. Capacity of old trees to respond to environmental change. *J. Integr. Plant Biol.* **50**, 1355–1364 (2008).
15. Piper, F. I. & Fajardo, A. No evidence of carbon limitation with tree age and height in *Nothofagus pumilio* under Mediterranean and temperate climate conditions. *Ann. Bot.* **108**, 907–917 (2011).
16. Weiner, J. & Thomas, S. C. The nature of tree growth and the “age-related decline in forest productivity”. *Oikos* **94**, 374–376 (2001).
17. Jenkins, J. C., Chojnacky, D. C., Heath, L. S. & Birdsey, R. A. *Comprehensive Database of Diameter-based Biomass Regressions for North American Tree Species* General Technical Report NE-319, <http://www.nrs.fs.fed.us/pubs/6725> (USDA Forest Service, Northeastern Research Station, 2004).
18. Niklas, K. J. & Enquist, B. J. Canonical rules for plant organ biomass partitioning and annual allocation. *Am. J. Bot.* **89**, 812–819 (2002).
19. Thomas, S. C. Photosynthetic capacity peaks at intermediate size in temperate deciduous trees. *Tree Physiol.* **30**, 555–573 (2010).
20. Steppe, K., Niinemets, Ü. & Teskey, R. O. in *Size- and Age-Related Changes in Tree Structure and Function* (eds Meinzer, F. C., Lachenbruch, B. & Dawson, T. E.) 235–253 (Springer, 2011).
21. Gilmore, D. W. & Seymour, R. S. Alternative measures of stem growth efficiency applied to *Abies balsamea* from four canopy positions in central Maine, USA. *For. Ecol. Manage.* **84**, 209–218 (1996).
22. Kaufmann, M. R. & Ryan, M. G. Physiographic, stand, and environmental effects on individual tree growth and growth efficiency in subalpine forests. *Tree Physiol.* **2**, 47–59 (1986).
23. Coomes, D. A., Holdaway, R. J., Kobe, R. K., Lines, E. R. & Allen, R. B. A general integrative framework for modelling woody biomass production and carbon sequestration rates in forests. *J. Ecol.* **100**, 42–64 (2012).
24. Binkley, D. A hypothesis about the interaction of tree dominance and stand production through stand development. *For. Ecol. Manage.* **190**, 265–271 (2004).
25. Pretzsch, H. & Biber, P. A re-evaluation of Reineke’s rule and stand density index. *For. Sci.* **51**, 304–320 (2005).
26. Kashian, D. M., Turner, M. G., Romme, W. H. & Lorimer, C. G. Variability and convergence in stand structural development on a fire-dominated subalpine landscape. *Ecology* **86**, 643–654 (2005).
27. Munné-Bosch, S. Do perennials really senesce? *Trends Plant Sci.* **13**, 216–220 (2008).
28. Jump, A. S., Hunt, J. M. & Peñuelas, J. Rapid climate change-related growth decline at the southern range edge of *Fagus sylvatica*. *Glob. Change Biol.* **12**, 2163–2174 (2006).
29. Lindenmayer, D. B., Laurance, W. F. & Franklin, J. F. Global decline in large old trees. *Science* **338**, 1305–1306 (2012).
30. Enquist, B. J., West, G. B. & Brown, J. H. Extensions and evaluations of a general quantitative theory of forest structure and dynamics. *Proc. Natl Acad. Sci. USA* **106**, 7046–7051 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the hundreds of people who have established and maintained the forest plots and their associated databases; M. G. Ryan for comments on the manuscript; C. D. Canham and T. Hart for supplying data; C. D. Canham for discussions and feedback; J. S. Baron for hosting our workshops; and Spain’s Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA) for granting access to the Spanish Forest Inventory Data. Our analyses were supported by the United States Geological Survey (USGS) John Wesley Powell Center for Analysis and Synthesis, the USGS Ecosystems and Climate and Land Use Change mission areas, the Smithsonian Institution Global Earth Observatory—Center for Tropical Forest Science (CTFS), and a University of Nebraska-Lincoln Program of Excellence in Population Biology Postdoctoral Fellowship (to N.G.B.). In addition, X.W. was supported by National Natural Science Foundation of China (31370444) and State Key Laboratory of Forest and Soil Ecology (LFSE2013-11). Data collection was funded by a broad range of organizations including the USGS, the CTFS, the US National Science Foundation, the Andrews LTER (NSF-LTER DEB-0823380), the US National Park Service, the US Forest Service (USFS), the USFS Forest Inventory and Analysis Program, the John D. and Catherine T. MacArthur Foundation, the Andrew W. Mellon Foundation, MAGRAMA, the Council of Agriculture of Taiwan, the National Science Council of Taiwan, the National Natural Science Foundation of China, the Knowledge Innovation Program of the Chinese Academy of Sciences, Landcare Research and the National Vegetation Survey Database (NVS) of New Zealand, the French Fund for the Global Environment and Fundación ProYungas. This paper is a contribution from the Western Mountain Initiative, a USGS global change research project. Any use of trade names is for descriptive purposes only and does not imply endorsement by the USA government.

Author Contributions N.L.S. and A.J.D. conceived the study with feedback from R.C. and D.A.C., N.L.S., A.J.D., R.C. and S.E.R. wrote the manuscript. R.C. devised the main analytical approach and wrote the computer code. N.L.S., A.J.D., R.C., S.E.R., P.J.B., N.G.B., D.A.C., E.R.L., W.K.M. and N.R. performed analyses. N.L.S., A.J.D., R.C., S.E.R., P.J.B., D.A.C., E.R.L., W.K.M., E.Á., C.B., S.B., G.C., S.J.D., A.D., C.N.E., O.F., J.F.F., H.R.G., Z.H., M.E.H., S.P.H., D.K., Y.L., J.-R.M., A.M., L.R.M., R.J.P., N.P., S.-H.S., I.-F.S., S.T., D.T., P.J.v.M., X.W., S.K.W. and M.A.Z. supplied data and sources of allometric equations appropriate to their data.

Author Information Fitted model parameters for each species have been deposited in USGS’s ScienceBase at <http://dx.doi.org/10.5066/F7JS9NFM>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.L.S. (nstephenson@usgs.gov).

Sensory–motor transformations for speech occur bilaterally

Gregory B. Cogan¹, Thomas Thesen², Chad Carlson^{2†}, Werner Doyle³, Orrin Devinsky^{2,3} & Bijan Pesaran¹

Historically, the study of speech processing has emphasized a strong link between auditory perceptual input and motor production output^{1–4}. A kind of ‘parity’ is essential, as both perception- and production-based representations must form a unified interface to facilitate access to higher-order language processes such as syntax and semantics, believed to be computed in the dominant, typically left hemisphere^{5,6}. Although various theories have been proposed to unite perception and production^{2,7}, the underlying neural mechanisms are unclear. Early models of speech and language processing proposed that perceptual processing occurred in the left posterior superior temporal gyrus (Wernicke’s area) and motor production processes occurred in the left inferior frontal gyrus (Broca’s area)^{8,9}. Sensory activity was proposed to link to production activity through connecting fibre tracts, forming the left lateralized speech sensory–motor system¹⁰. Although recent evidence indicates that speech perception occurs bilaterally^{11–13}, prevailing models maintain that the speech sensory–motor system is left lateralized^{11,14–18} and facilitates the transformation from sensory-based auditory representations to motor-based production representations^{11,15,16}. However, evidence for the lateralized computation of sensory–motor speech transformations is indirect and primarily comes from stroke patients that have speech repetition deficits (conduction aphasia) and studies using covert speech and haemodynamic functional imaging^{16,19}. Whether the speech sensory–motor system is lateralized, like higher-order language processes, or bilateral, like speech perception, is controversial. Here we use direct neural recordings in subjects performing sensory–motor tasks involving overt speech production to show that sensory–motor transformations occur bilaterally. We demonstrate that electrodes over bilateral inferior frontal, inferior parietal, superior temporal, premotor and somatosensory cortices exhibit robust sensory–motor neural responses during both perception and production in an overt word-repetition task. Using a non-word transformation task, we show that bilateral sensory–motor responses can perform transformations between speech-perception- and speech-production-based representations. These results establish a bilateral sublexical speech sensory–motor system.

To investigate the sensory–motor representations that link speech perception and production, we used electrocorticography (ECoG), in which electrical recordings of neural activity are made directly from the cortical surface in a group of patients with pharmacologically intractable epilepsy. ECoG is an important electrophysiological signal recording modality that combines excellent temporal resolution with good spatial localization. Critically for this study, ECoG data contain limited artefacts due to muscle and movements during speech production compared with non-invasive methods that suffer artefacts with jaw movement²⁰. Thus, using ECoG we were able to investigate directly neural representations for sensory–motor transformations using overt speech production.

Sixteen patients with subdural electrodes (see Supplementary Figs 1 and 2) implanted in the left hemisphere (6 subjects), right hemisphere

(7 subjects) or both hemispheres (3 subjects) performed variants of an overt word repetition task designed to elicit sensory–motor activations (Fig. 1a, Methods and Supplementary Table 1). We observed increases in neural activity across the high gamma frequency range (60–200 Hz and above) with maximal activity across subjects between 70–90 Hz. High gamma activity reflects the spiking activity of populations of neurons during task performance^{20,21}. Individual electrodes showed one of three types of task responses: sensory–motor (S-M), production (PROD), or auditory (AUD) (Fig. 1b, see Methods). We found that AUD activity was generally localized to the superior temporal gyrus and middle temporal gyrus (42 out of 57 electrodes (74%); Fig. 2a, b) and PROD activity occurred mostly in the motor and premotor cortices, somatosensory cortex, and the inferior parietal lobule (98 out of 124 electrodes (79%); Fig. 2a, b), consistent with previous models and results of speech-perception and -production studies^{11,12,17}. Furthermore, electrical stimulation of PROD electrode locations resulted in orofacial movements consistent with a motor function (see Supplementary

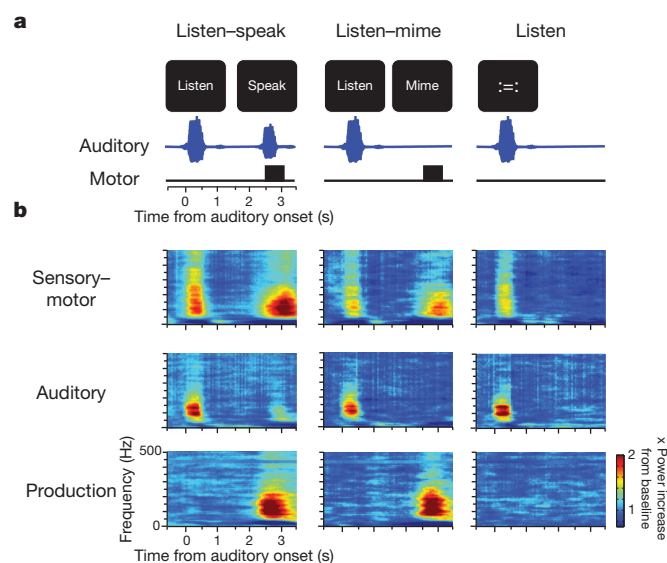


Figure 1 | Behavioural tasks and example neural activations. **a**, 16 Subjects were presented with an auditory consonant–vowel–consonant single-syllable word and instructed to perform one of three tasks on interleaved trials: listen–speak (listen to the word, visual prompt ‘Listen’, then after a 2-s delay repeat the word, visual prompt ‘Speak’); listen–mime (listen to the word, visual prompt ‘Listen’, then after a 2-s delay, mime speaking the word, visual prompt ‘Mime’); listen (passively listen to the word, visual prompt ‘=:’). Auditory and motor timelines are shown. **b**, Example time–frequency spectrograms of ECoG activity normalized at each frequency to the baseline power during visual prompt. AUD, significant activity during each task epoch with auditory stimuli; PROD, significant activity during both production epochs; S-M, significant activity during the auditory and production epochs in listen–speak and listen–mime tasks.

¹Center for Neural Science, New York University, New York, New York 10003, USA. ²Department of Neurology, New York University School of Medicine, New York, New York 10016, USA. ³Department of Neurosurgery, New York University School of Medicine, New York, New York 10016, USA. [†]Present address: Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA.

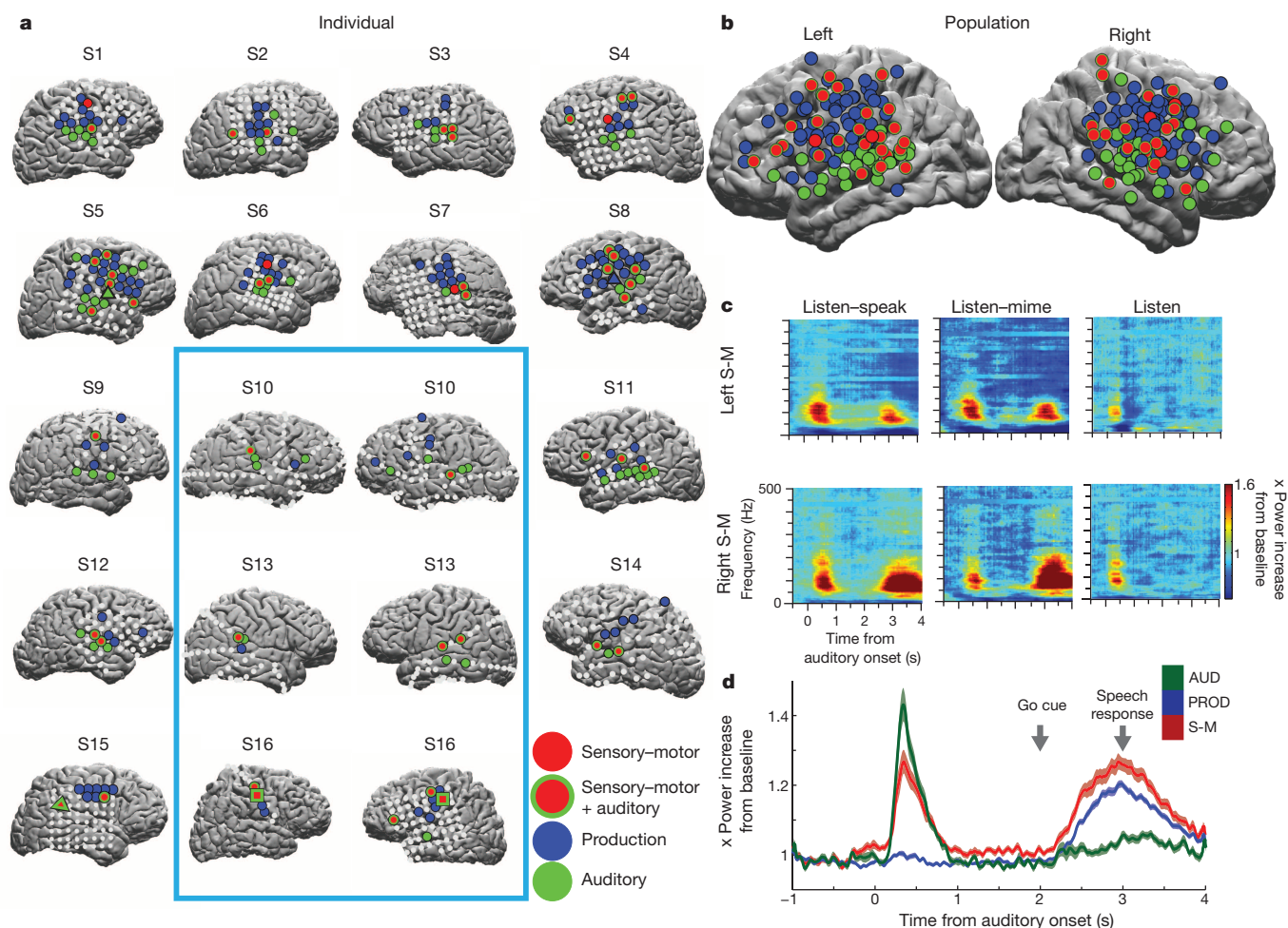


Figure 2 | Topography of neural responses and bilateral activation.

a, Significant task-related activations within individual subject brains for left (subjects S3, S4, S7, S8, S11, S14), right (S1, S2, S5, S6, S9, S12, S15), or both (S10, S13, S16) hemispheres. Bilateral coverage is indicated by the light blue box. Electrodes with significant high gamma activity (70–90 Hz) are shown for AUD (green), PROD (blue) and S-M (red) activations. AUD and S-M activations (red with green) were often present on the same electrode. Electrodes without significant activation are shown in grey. Triangles denote example activations from Fig. 1b, and squares (S16) denote example

spectrograms in Fig. 2c. **b**, Significant electrodes projected onto population average left and right hemispheres, colours as in **a**. Electrode sizes have been increased for illustrative purposes (for actual sizes see Supplementary Fig. 4). Neural spectrograms for example S-M electrodes in left and right hemispheres of S16 during listen-speak, listen-mime and listen tasks. **d**, Population average neural response profiles for each class of electrodes. Shaded regions indicate s.e.m. values across electrodes. Go cue and average production response onset are indicated by grey arrows.

Fig. 3). Critically, contrary to one of the core dogmas of brain and language, S-M activity occurred bilaterally in the supramarginal gyrus, middle temporal gyrus, superior temporal gyrus, somatosensory cortex, motor cortex, premotor cortex and inferior frontal gyrus (Fig. 2a, b, 49 electrodes; see Supplementary Table 2 and Supplementary Fig. 4) and was observed in all subjects (Fig. 2a). Of the 49 S-M sites, 45 sites showed auditory activation during the 'listen task' (Fig. 2a, b; Supplementary Figs 4 and 5; 45 out of 49 electrodes (approximately 92%)), suggesting a role in speech perception. Hemispheric dominance as determined by Wada testing did not correlate with the hemisphere of the electrode placement ($\chi^2(3) = 0.92$, $P = 0.34$). Importantly, in three subjects with bilateral coverage, S-M activity was present on electrodes in both hemispheres (Fig. 2a, c) and the likelihood of an electrode being a S-M site did not differ between hemispheres (Fisher's exact test, $P = 0.31$). These results demonstrate that S-M activity occurs bilaterally.

Given the evidence for bilateral S-M activity, we performed a series of analyses and experimental manipulations to test the hypothesis that bilateral S-M activity is in fact sensory-motor and represents sensory-motor transformations for speech.

One concern is that S-M activity is not due to sensory and motor processes but to sensory activation in both auditory (input) and production epochs (sound of your own voice). We observed several convergent

lines of evidence that S-M activity reflects both sensory and motor processing (see Fig. 2d and Methods). First, S-M sites contain a sensory response because they responded to auditory stimulation as rapidly as AUD sites (S-M latency = 158 ms, AUD = 164 ms; see Fig. 2d). Second, S-M responses during production are not due to auditory sensory reafferent input from hearing one's own voice because responses were present during the 'listen-mime task' as well as the 'listen-speak task'. Third, S-M responses during production are not due to somatosensory reafference from moving articulators because S-M activity significantly increased within 248 ms of the production 'go' cue, whereas vocal responses occurred substantially later at 1,002 ms (± 40 ms s.e.m.). Fourth, S-M production responses contain motor features because they occurred together with, and even before, PROD electrode responses (S-M = 248 ms, PROD = 302 ms, $Q = 0.03$; permutation test; see Methods). Finally, S-M activity was persistently elevated during the delay period ($P = 0.01$; see Fig. 2d, Methods), broadly consistent with planning activity, unlike PROD delay-period activity ($P = 0.64$) or AUD delay-period activity ($P = 0.53$). These results demonstrate that S-M activity cannot be simply sensory and spans both sensory and motor processes.

A related concern is that sensory-motor transformations are first carried out in the left hemisphere. If so, S-M responses in the right hemisphere could be due to communication from the left hemisphere. To test this

hypothesis, we further examined latencies of S-M responses according to hemisphere. Response latencies did not differ significantly in each hemisphere in either the auditory (right hemisphere, 156 ms; left hemisphere, 182 ms; $Q = 1 \times 10^{-4}$; permutation test) or the production epoch (right hemisphere, 272 ms; left hemisphere, 268 ms; $Q = 1 \times 10^{-4}$; see Methods). Therefore, right hemisphere responses cannot be due to computations that were first carried out in the left hemisphere and the data do not support strictly lateralized sensory-motor computations.

Another concern is that S-M activity may not reflect speech processing and may also be present during simple auditory inputs and orofacial motor outputs. To test this, we employed a 'tone-move task' in one of the bilaterally implanted subjects (subject 13 (S13); see Methods). We found that S-M electrodes did not have significant sensory-motor responses during the tone-move task ($P = 0.36$, permutation test; see Supplementary Fig. 6). Thus, S-M activity is specific to mapping sounds to goal-directed vocal motor acts and is likely specific to speech (see Supplementary Discussion 1.3).

Thus far we have shown the S-M activity is bilateral, sensory-motor, and likely to be specific to speech. However, an important open question is whether S-M responses reflect the transformation that links speech perception and production and can support a unified perception-production representational interface. A specific concern is that high gamma ECoG activity may pool heterogeneous neural responses beneath the electrode. S-M responses may combine activity from neurons which encode perceptual processes active during the auditory cue and other neurons which encode production processes active during the utterance. If this is true, none of the activity necessarily reflects a sensory-motor transformation that links perception and production. To be able to rule out this alternative and demonstrate that S-M responses are involved in sensory-motor transformations, we reasoned that two requirements must be met. S-M activity must encode information about the content of the underlying speech processes, and this encoding must reflect transformative coding between the sensory input and motor output.

To test whether S-M activity encodes information about speech content, we decoded the neural activity to predict, on each trial, what the subjects heard and said. We used seven consonant-vowel-consonant words (heat, hit, hat, hoot, het, hot and hut) and trained a seven-way linear classifier to decode the neural responses (see Methods). Individual electrodes only weakly encoded speech content, but when we decoded activity pooled across groups of electrodes, we found that all three electrode groups encoded speech tokens (see Fig. 3). AUD electrodes performed best with an average classification performance of 42.7% ($\chi^2(1) = 56.5$, $P = 6 \times 10^{-14}$), followed by S-M electrodes, which showed performance of 33.4% ($\chi^2(1) = 25.6$, $P = 4 \times 10^{-7}$), and then PROD electrodes, which showed performance of 27.1% ($\chi^2(1) = 11.5$, $P = 7 \times 10^{-4}$). Furthermore, classification performance for S-M electrodes did not differ between the two hemispheres (left hemisphere, 29%; right hemisphere, 27%; Fisher's exact test, $P = 0.5$; Fig. 3c). Thus, bilateral S-M activity encodes information about the sensory and motor contents of speech, meeting an important requirement for sensory-motor transformations.

We next sought to test whether S-M activity can link speech perception and production by transforming auditory input into production output. The essential requirement for transformation is that neural encoding of sensory input should depend on subsequent motor output. Previous work has characterized visual-motor transformations using a transformation task in which the spatial location of a visual cue can instruct a motor response to the same or different spatial location (the 'pro-anti task')^{22,23}. Sensory-motor neurons in the dorsal visual stream display different responses to the visual cue depending on the motor contingency, demonstrating a role for these neurons in the visual-motor transformation²².

Given these predictions from animal neurophysiology, we tested four subjects as they performed an auditory-motor transformation task (the listen-speak transformation task) that employed two non-words (kig, pob) to examine whether S-M activity has a role in transformations for

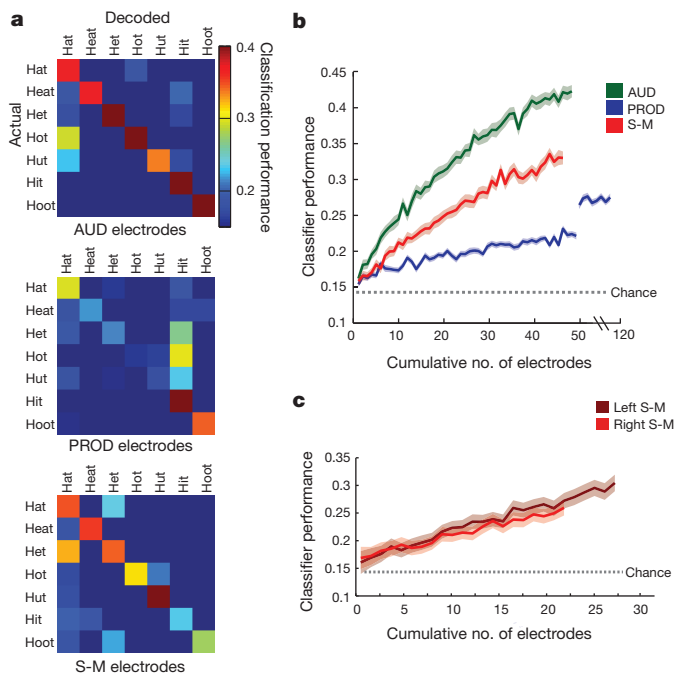


Figure 3 | Neural decoding of words. **a**, Confusion matrices show proportion identified for a seven-way linear classifier using neural responses. AUD electrodes (top), PROD electrodes (middle) and S-M electrodes (bottom) are shown. The threshold for performance is at chance level, $P = 0.14$, for the purposes of displaying the electrodes clearly. **b**, Classification performance for increasing numbers of electrodes. Chance performance is indicated by the dotted line. **c**, Classification performance for S-M electrodes in the left and right hemispheres. Methods present S-M results by response epoch.

speech (see Fig. 4A, Supplementary Figs 7 and 8 and Methods). This task enabled us to hold the sensory and motor components constant while manipulating the transformation process itself in order to measure how the encoding of this content changed depending on how perceptual input was mapped onto production output. The use of non-words instead of words offered other advantages. Non-words enabled us to examine sublexical transformations for speech and could be designed to differ maximally in their articulatory dimensions and their neural representations (see Methods and Supplementary Discussion 1.1 and 1.2).

At least three models describe how neural responses encode the task variables. If responses follow a strictly sensory model, the encoding will follow the content of the sensory inputs and confuse trial conditions in which kig is converted to kig (kig→kig) with trials in which kig is converted to pob (kig→pob), as well as trials in which pob is converted to pob (pob→pob) with trials in which pob is converted to kig (pob→kig) (see Fig. 4Ba). Conversely, responses that follow a strictly motor model will encode the production outputs, confusing kig→kig with pob→kig trials and pob→pob with kig→pob trials (see Fig. 4Bb). If S-M responses pool responses from sensory and motor neurons, the encoding will follow the sensory model during sensory input and the motor model during motor output. In contrast, S-M responses that reflect the transformation of sensory input into motor output must follow a different transformation model and encode the sensory information differently depending on the upcoming motor act (see Fig. 4Bc). Neural activity displaying this property could compute a representational transformation (see Supplementary Discussion 1.1, 1.2). If so, responses that follow a transformation model will not confuse trial conditions with either identical input or identical output. Consequently, each of the three models predicted very different patterns of neural coding.

We constructed linear classifiers to decode neural responses. As expected, AUD electrodes in the auditory epoch encoded the auditory input (Fig. 4Ba, Ca) and PROD electrodes encoded the output during

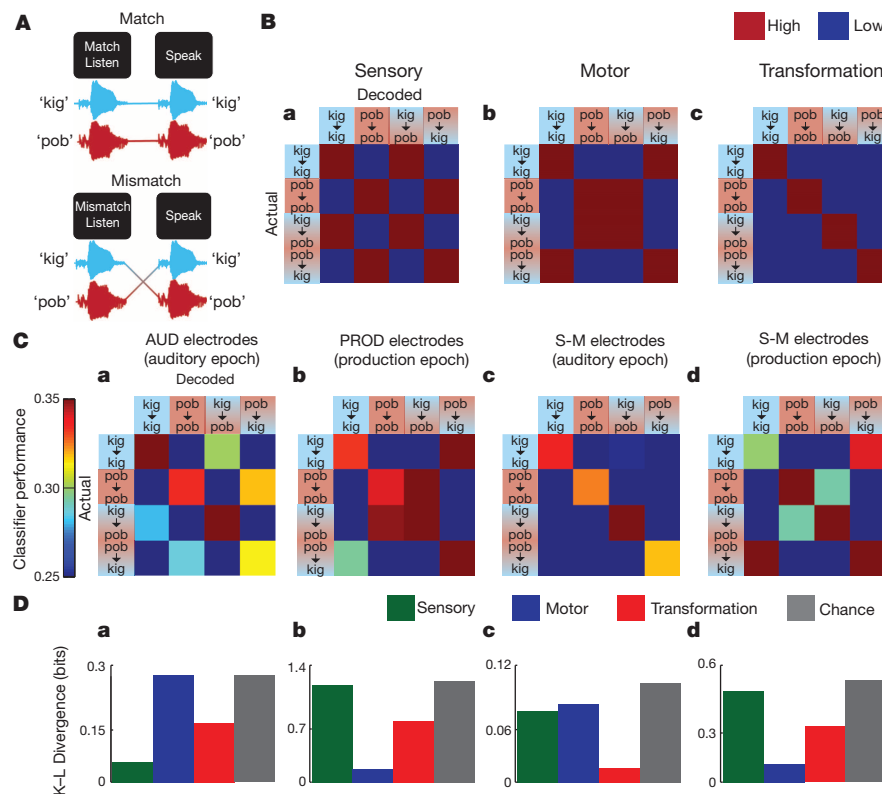


Figure 4 | Listen-speak transformation task. **A**, In the listen-speak transformation task, subjects have to transform a non-word they hear into a non-word they speak according to a simple rule. Subjects were first presented with a visual cue: 'Match Listen' or 'Mismatch Listen' that instructed the rule that determined the non-word to say in response to the non-word they heard. On 'match trials' the rule was to repeat the non-word they heard. On 'mismatch trials', the rule was to say the non-word that they did not hear. The non-words were 'kig' and 'pob'. Subjects then heard one of the

two non-words, waited for a short delay, then said the appropriate non-word in response to the 'Speak' cue. There were four task conditions: kig→kig (hear 'kig' and say 'kig'); pob→pob (hear 'pob' and say 'pob'); kig→pob (hear 'kig' and say 'pob'); and pob→kig (hear 'pob' and say 'kig'). **B**, **a–c**, Confusion matrices predicted by the sensory, motor and transformation models with high and low classification scores. **C**, **a–d**, Confusion matrices during the listen-speak transformation task. **D**, **a–d**, Model fit quantified using a Kullback–Leibler (K–L) divergence.

the production epoch (utterance; Fig. 4Bb, Cb). However, S-M electrodes changed their encoding over the course of the trial. During the auditory epoch, S-M electrodes encoded both sensory and motor conditions concurrently, consistent with the presence of a sensory-motor transformation (Fig. 4Bc, Cc). Interestingly, during the production epoch, S-M responses no longer encoded the auditory input and encoded the production output (Fig. 4Cd), suggesting the transformation has largely been computed by that time. To quantify the comparison of different models, we used the Kullback–Leibler divergence (see Fig. 4Da–d, Methods). The results were consistent with the response patterns in the confusion matrices.

We can also rule out that the difference in S-M responses is due to a third population of neurons that selectively responds to the cue instructing how perceptual input was mapped onto production output ('match' or 'mismatch'). We ran the same linear classifier during cue presentation and found that the S-M responses did not encode the cue ($\chi^2(1) = 0.08$, $P = 0.78$; see Methods).

Using direct brain recordings (ECoG) and overt speech, we demonstrate that a sensory-motor system for transforming sublexical speech signals exists bilaterally. Our results are in keeping with models of speech perception that posit bilateral processing but contradict models that posit lateralized sensory-motor transformations^{11,16}. Our results also highlight how S-M activity during perceptual input reflects the transformation of speech sensory input into motor output. We propose that the presence of such transformative activity demonstrates a unified sensory-motor representational interface that links speech-perception- and speech-production-based representations. Such an interface is important during speech articulation, acquisition and self-monitoring^{24–26}. As

right hemisphere lesions do not give rise to conduction aphasia^{19,27–29}, our evidence for bilateral sensory-motor transformations promotes an interesting distinction between speech and language: although sensory-motor transformations are bilateral, the computational system for higher-order language is lateralized^{5,6} (see Supplementary Fig. 9). This hypothesis invokes a strong interface between sensory-based speech-perception representations and motor-based speech-production representations and suggests that deficits for conduction aphasia are more abstract and linguistic in nature. We propose that bilateral sublexical transformations could support a unification of perception- and production-based representations into a sensory-motor interface⁶, drawing a distinction between the bilateral perception-production functions of speech and lateralized higher order language processes.

METHODS SUMMARY

Electrocorticographic (ECoG) recordings were obtained from 16 patients (10 females) undergoing treatment for pharmacologically resistant epilepsy. Each patient provided informed consent in accordance with the Institutional Review Board at New York University Langone Medical Center. Grid implantation was in the left hemisphere (6 subjects), right hemisphere (7 subjects) or both hemispheres (3 subjects). All 16 subjects performed an overt word repetition task (listen-speak task) as well as two control tasks (listen-mime task³⁰ and listen task). One subject also performed a tone-move task. Four subjects also performed a listen-speak transformation task involving non-words. ECoG recordings were made using both grid and strip electrode arrays with 2.3-mm contact size and 10-mm spacing. Spectral analysis was performed using 500-ms analysis windows with ± 5 -Hz frequency smoothing and a stepping size of 50 ms. Neural responses were defined as high gamma neural activity between 70 and 90 Hz and significance was assessed using a shuffling procedure. Classification analyses were carried out using a linear discriminant analysis with high gamma power spectral features.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 September; accepted 2 December 2013.

Published online 15 January 2014.

1. Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the Speech Code. *Psychol. Rev.* **74**, 431–461 (1967).
2. Liberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21**, 1–36 (1985).
3. Halle, M. & Stevens, K. N. Speech recognition: a model and a program for research. *IEEE Trans. Inf. Theory* **8**, 155–159 (1962).
4. Halle, M. & Stevens, K. N. Analysis by synthesis In (eds Wathen-Dunne, W. & Woods, L. E.) *Proceedings of seminar on speech compression and processing* Vol. 2 paper D7 1959).
5. Berwick, R. C., Friederici, A. D., Chomsky, N. & Bolhuis, J. J. Evolution, brain, and the nature of language. *Trends Cogn. Sci.* **17**, 89–98 (2013).
6. Chomsky, N. *The Minimalist Program* (MIT Press, 1995).
7. Jakobson, R. *Child Language, Aphasia and Phonological Universals* (Mouton, 1968).
8. Lichtheim, L. On aphasia. *Brain* **7**, 433–484 (1885).
9. Wernicke, C. The aphasic symptom-complex: a psychological study on an anatomical basis. *Arch. Neurol.* **22**, 280–282 (1970).
10. Geschwind, N. Disconnexion syndromes in animals and man. I. *Brain* **88**, 237–294 (1965).
11. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nature Rev. Neurosci.* **8**, 393–402 (2007).
12. Price, C. J. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. NY Acad. Sci.* **1191**, 62–88 (2010).
13. Obleser, J., Eisner, F. & Kotz, S. A. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* **28**, 8116–8123 (2008).
14. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Rev. Neurosci.* **12**, 718–724 (2009).
15. Hickok, G., Houde, J. & Rong, F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* **69**, 407–422 (2011).
16. Hickok, G., Okada, K. & Serences, J. T. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* **101**, 2725–2732 (2009).
17. Guenther, F. H. Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* **39**, 350–365 (2006).
18. Wise, R. J. S. *et al.* Separate neural subsystems within 'Wernicke's area'. *Brain* **124**, 83–95 (2001).
19. Caramazza, A., Basili, A. G., Koller, J. J. & Berndt, R. S. An investigation of repetition and language processing in a case of conduction aphasia. *Brain Lang.* **14**, 235–271 (1981).
20. Crone, N. E., Sinai, A. & Korzeniewska, A. Event-related dynamics of brain oscillations. *Prog. Brain Res.* **159**, 275–295 (2006).
21. Markowitz, D. A., Wong, Y. T., Gray, C. M. & Pesaran, B. Optimizing the decoding of movement goals from local field potentials in macaque cortex. *J. Neurosci.* **31**, 18412–18422 (2011).
22. Zhang, M. & Barash, S. Neuronal switching of sensorimotor transformations for antisaccades. *Nature* **408**, 971–975 (2000).
23. Gail, A. & Andersen, R. Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. *J. Neurosci.* **26**, 9376–9384 (2006).
24. Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S. & Houde, J. F. Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc. Natl Acad. Sci. USA* **110**, 2653–2658 (2013).
25. Oller, D. K., Eilers, R. E. & Oiler, D. K. The role of audition in infant babbling the role of audition in infant babbling. *Child Dev.* **59**, 441–449 (1988).
26. Agnew, Z. K., McGettigan, C., Banks, B. & Scott, S. K. Articulatory movements modulate auditory responses to speech. *Neuroimage* **73**, 191–199 (2013).
27. Goodglass, H., Kaplan, E. & Barresi, B. *Assessment of Aphasia and Related Disorders* (Lippincott Williams & Wilkins, 2000).
28. Damasio, H. & Damasio, A. R. The anatomical basis of conduction aphasia. *Brain* **103**, 337–350 (1980).
29. Benson, D. F. *et al.* Conduction aphasia: a clinicopathic study. *Arch. Neurol.* **28**, 339–346 (1973).
30. Murphy, K. *et al.* Cerebral areas associated with motor control of speech in humans cerebral areas associated with motor control of speech in humans. *J. Appl. Physiol.* **83**, 1438–1447 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank A. Weiss, J. MacArthur and L. Frank for developing the data acquisition hardware and software; O. Felsovalyi, E. Londen, P. Purushothaman, L. Melloni, C. Boomhaur and A. Trongnetrpunya for technical assistance; and D. Poeppel and C. Brody for comments on the manuscript. This work was supported, in part, by R03-DC010475 from the NIDCD, a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund (B.P.), a Watson Investigator Program Award from NYSTAR (B.P.), a McKnight Scholar Award (B.P.) and a Sloan Research Fellowship (B.P.).

Author Contributions G.B.C. designed the experiment, performed the research, analysed the data and wrote the manuscript. T.T. and O.D. performed the research and wrote the manuscript. C.C. and W.D. performed the research. B.P. designed the experiment, performed the research, analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.P. (bijan@nyu.edu).

METHODS

Participants. Electrocorticographic (ECoG) recordings were obtained from 16 patients (6 males, 10 females; see Supplementary Table 1) with pharmacologically resistant epilepsy undergoing clinically motivated subdural electrode recordings at the New York University School of Medicine Comprehensive Epilepsy Center. Informed Consent was obtained from each patient in accordance with the Institutional Review Board at the New York University Langone Medical Center. Patient selection for the present study followed strict criteria: first, cognitive and language abilities in the average range or above, including language and reading ability, as indicated by formal neuropsychological testing (see Supplementary Table 1); and second, normal language organization as indicated by cortical stimulation mapping, when available. In addition, only electrode contacts outside the seizure onset zone and with normal interictal activity were included in the analysis.

Behavioural tasks and recordings. All participants performed three behavioural tasks: listen–speak, listen–mime and listen (Fig. 1a). Behavioural tasks were performed while participants reclined in a hospital bed. Tasks were controlled by a computer placed on the service tray on the bed running the Presentation program (NeuroBehavioural Systems). Behavioural audio recordings were either synchronized with the neural recordings at 10 kHz (see below) or recorded on the computer and referenced to the go cue. For a subset of subjects, a video camera with built-in microphone (Sony) was positioned to monitor subject orofacial movements and utterances. Video was streamed to disk (Adobe Premier Pro. (video at 29.95 frames per s, and audio at 44.1 kHz)). Audio-visual and neural signals were synchronized video-frame-by-video-frame using an Analogue-to-Digital Video Converter (Canopus).

Listen–speak, listen–mime³⁰ and listen tasks were randomly interleaved on a trial-by-trial basis with at least 4 s between trials. Each trial began with a visual cue presented, followed by the auditory consonant–vowel–consonant (CVC) token 500 ms later. We used CVC words composed of the same consonants, ‘h’ and ‘t’, and different vowels (hat, hit, heat, hoot, het, hut, hot). These spoken syllables span the vowel space and differ in their auditory and articulatory content. Subjects had to either listen passively (listen), repeat the syllable after a cue (listen–speak) or mime the syllable after a different cue (listen–mime, produce the appropriate mouth movements but with no vocal cord vibration³⁰; see Supplementary Fig. 10). The temporal delay between the auditory cue and the movement cue was 2 s. We obtained between 49 and 166 trials per condition (within subject) and between 175 and 334 total trials per subject.

For the tone–move task (see Supplementary Fig. 6), after the listen cue was delivered, a 500-ms, 1,000-Hz sinusoidal tone (with 100-ms on and off ramps) was presented. After a short, 2-s delay another visual cue was presented (move) instructing the subject to move their articulators (tongue, lips and jaw). For one subject, these trials were randomly interleaved within blocks of the listen–speak, listen–mime and listen tasks (see above).

For the listen–speak transformation task, four subjects (see Supplementary Figs 7 and 8) were first presented with one of two visual cues: ‘Match Listen’ or ‘Mismatch Listen’. After a delay, subjects heard one of two non-words: ‘kig’ (/kɪg/) or ‘pob’ (/pɒb/). These non-words were chosen to differ maximally on their articulator dimensions: ‘kig’ contains a velar (back) voiceless stop consonant, followed by a high front vowel and finally a velar voiced stop consonant, and ‘pob’ contains a bilabial (front) voiceless stop consonant followed by a back low vowel and then a bilabial front voiced stop consonant. The tongue movement therefore goes back to front to back for ‘kig’ and front to back to front for ‘pob’. The reason for choosing maximally different articulations was that larger articulator differences might lead to larger neural activity differences. After a short delay (randomized between 1.5 and 2 s), another visual cue was presented (‘Speak’) to which subjects were to respond by saying the match non-word they had heard if they had seen the initial match cue, or say the mismatch non-word if they had seen the mismatch cue. Each non-word within each condition was presented between 63 and 78 times per subject, with total trials ranging from 255 to 309 per subject. This control was carried out in separate blocks trials that alternated with blocks of the main listen–speak, listen–mime and listen tasks.

Surface reconstruction and electrode localization. To localize electrode recording sites, pre-surgical and post-surgical T1-weighted magnetic resonance imaging (MRI) scans were obtained for each patient and co-registered with each other³¹. The co-registered images were then normalized to an MNI-152 template and electrode locations were then extracted in MNI (Montreal Neurological Institute) space (projected to the surface) using the co-registered image, followed by skull stripping³². A three-dimensional reconstruction of each patient’s brain was computed using FreeSurfer (Fig. 2; S2, S3, S4, S5, S6, S7, S8 and S10 (ref. 33)). For Supplementary Table 2, Talairach coordinates were converted from MNI space using the EEG/MRI toolbox in Matlab (<http://sourceforge.net/projects/eeeg/>, GNU General Public License).

Neural recordings and preprocessing. EEG data were recorded from intracranially implanted subdural electrodes (AdTech Medical Instrument Corp.) in patients undergoing elective monitoring of pharmacologically intractable seizures. Electrode placement was based entirely on clinical grounds for identification of seizure foci and eloquent cortex during stimulation mapping, and included grid (8 × 8 contacts), depth (1 × 8 contacts) and strip (1 × 4 to 1 × 12 contacts) electrode arrays with 10-mm inter-electrode spacing centre-to-centre. Subdural stainless steel recording grid and strip contacts were 4 mm in diameter; consequently the distance between contacts was 6 mm and they had an exposed 2.3-mm diameter recording contact.

For 7 of the 16 subjects, neural signals from up to 256 channels were amplified (×10, INA121 Burr-Brown instrumentation amplifier), bandpass filtered between 0.1–4,000 Hz and digitized at 10 kHz (NSpike, Harvard Instrumentation Laboratories) before being continuously streamed to disk for off-line analysis (custom C and Matlab code). The front-end amplifier system was powered by sealed lead acid batteries (Powersonic) and optically isolated from the subject. After acquisition, neuronal recordings were further low-pass filtered at 800 Hz and down-sampled offline to 2,000 Hz for all subsequent analysis. For the remaining 9 subjects, neural signals from up to 128 channels were recorded on a Nicolet One EEG system, bandpass-filtered between 0.5–250 Hz and digitized at 512 Hz. In some recordings, modest electrical noise was removed using line-filters centred on 60, 120 and 180 Hz (ref. 34).

Data Analysis. For activation analysis, time-frequency decomposition was performed using a multi-taper spectral analysis³⁴. The power spectrum was calculated during a 500-ms analysis window with ± 5 Hz frequency smoothing stepped 50 ms between estimates. Single trials were removed from the analysis if the raw voltage exceeded eight standard deviations from the across trial pool, and noisy channels were removed from the analysis by visual inspection or if they did not contain at least 60% of the total trials after the standard deviation threshold removal.

Sensory–motor transformations were defined as activity in the gamma range (70–90 Hz) that followed the auditory stimulus as well as the production cue during both listen–speak and listen–mime (Fig. 1b). As the example responses illustrate, some electrodes showed consistent increases in activity in the high gamma band as high as 300 Hz. As the frequency extent varied across subjects, we chose to focus on the 70–90-Hz frequency range as this band showed the greatest activation consistently across all subjects. Similar results were obtained when a broader frequency range extending up to 150 Hz was analysed. Although the listen–mime condition does involve altering the motor plan (no vocal cord vibration), sensory–motor activations were based on the conjunction of activity in both the listen–speak and the listen–mime conditions. Any neural activity that was specific to the listen–mime condition and not present in ‘normal’ speaking conditions was therefore excluded (see Supplementary Fig. 10).

Responses were divided into three types. The first response type, auditory (AUD), was defined as containing a response that was seen within 250–750 ms following the onset of the auditory stimulus in all three conditions (Fig. 1b, top). The second response type, production (PROD), was characterized as containing a response occurring between 500–1,000 ms after the respond cue in the listen–speak and the listen–mime conditions (Fig. 1b, middle). The last response type, S-M, contained both post stimulus and a post response cue activation in both the listen–speak and the listen–mime conditions (Fig. 1b bottom). The baseline period was defined as the 500 ms preceding the auditory stimulus.

In Fig. 1b, the experimental epoch was defined as –500 ms (pre) to 3,500 ms post auditory stimulus onset. In Fig. 2c the experimental epoch was defined as –500 ms (pre) to 4,000 ms post auditory stimulus onset. The additional 500 ms was included in Fig. 2c to compensate for slightly later production responses for that the subject. Power in each frequency band was normalized to the power in the baseline period by dividing by the power at each frequency. As the neural responses had variable onset times but were on average quite long in duration, the times were chosen to sample adequately all the responses under investigation.

To assess statistical significance, the average power across trials was taken in two time regions of interest for each trial within each condition. For the listen condition, the baseline values for each trial were shuffled with the post auditory values 10,000 times to create a null distribution. For the listen–speak and the listen–mime conditions, both the post-auditory and the post-production values were shuffled 10,000 times with the baseline values to create two null distributions. Initial significance was assessed using a permutation test by comparing the actual difference between the post auditory and post production values with the shuffled data differences³⁵. To correct for multiple comparisons, for all subjects, all three conditions and both analysis epochs (listen (post auditory), listen–speak (post auditory and post production) and listen–mime (post auditory and post production)) were pooled together and a false discovery rate (FDR) analysis was performed with an alpha threshold set at 0.05 (ref. 36).

The population latency analysis was performed using the baseline-corrected high gamma power response profiles for each electrode within each response class (S-M, AUD and PROD). The high gamma neural responses were first bandpass filtered (70–90 Hz) and then averaged within conditions. The listen–speak and listen–mime conditions were averaged together. As the data were recorded using two different sampling rates, the data were resampled to a 500 Hz sampling rate. To test for latencies within a response class, the latencies following either the auditory onset or the go cue were compared against the activity in the listen condition following the go cue by computing a permuted distribution for each time point. The significance values at each time point were then corrected for multiple comparisons using a FDR set with an alpha of 0.05. The first time point that was followed by at least 20 consecutive significant time points (40 ms) was taken to be the latency of the neural response. This resulted in four significant latency values. In the auditory epoch, AUD electrodes had significant neural responses at 164 ms and S-M electrodes had significant responses at 158 ms. During the production epoch, PROD electrodes had significant responses starting at 302 ms, whereas S-M electrodes had significant responses starting at 248 ms. A similar analysis was carried out comparing the left S-M electrodes with the right S-M electrodes, which resulted in four more significant latency values: right hemisphere (auditory 156 ms), left hemisphere (auditory 182 ms), right hemisphere (production 272 ms) and left hemisphere (production 268 ms). A direct comparison between these latencies within each task epoch using FDR-corrected shuffle tests (see above) revealed no significant results.

To assess whether or not during the auditory and production epochs, the S-M electrodes display significantly faster neural responses than the AUD and PROD electrodes, we repeated the permutation test, except instead of using the comparison of the task compared to the ‘Listen’ condition, we compared the S-M electrodes to the AUD electrodes in the auditory epoch and the S-M electrodes to the PROD electrodes in the production epoch. The results showed that whereas S-M and AUD electrodes did not differ in their latency values during the auditory epoch, S-M electrodes were significantly faster than PROD electrodes in the production epoch.

To test for power differences of the high gamma response (70–90 Hz) across hemispheres, we performed FDR-corrected permutation tests. Data were analysed by averaging a 300-ms time window, sliding 50 ms between estimates. The data were baseline-corrected (average –500 ms (pre) to 0 ms pre-stimulus activity across conditions, within electrodes) and then log-transformed before analysis. For each condition (listen–speak, listen–mime and listen) and within each hemisphere (left and right), we computed the task epoch responses by computing the average of the high gamma response during the auditory epoch (0–1,000 ms post auditory onset) and during the production epoch (0–1,500 ms post production cue onset). We then performed a series of permutation tests where we permuted the neural response across condition and/or across hemisphere, correcting for multiple comparisons using a FDR procedure. Only four tests produced significant results: listen–speak versus listen during the production epoch in each hemisphere, and listen–mime versus listen during the production epoch in each hemisphere. Furthermore, the neural responses within all conditions were not different across hemispheres (see Supplementary Fig. 11, $P > 0.05$, FDR corrected).

To assess the significant delay activation for each electrode class, a permutation test was carried out using filtered data as listed above. A permutation test was performed for each electrode class in which the average high gamma neural activity of the delay period (1–2 s post auditory onset) was compared to that of the baseline period (–1 s to –0.5 s pre auditory onset). Although PROD electrodes and AUD electrodes did not display elevated population neural activity ($P = 0.64$ and 0.53, respectively), S-M electrodes had significantly higher elevated delay activity compared to baseline ($P = 0.01$; see Fig. 2d).

Classification was performed using the single value decomposition (SVD) of the high gamma neural response (70–160-Hz, 300-ms sliding windows with an overlap of 250 ms) in either the auditory epoch (0–1,000 ms post auditory onset, AUD electrodes) or the production epoch (0–1,500 ms post go cue, PROD electrodes) or both (S-M electrodes). A linear discriminant analysis (LDA) classification was performed using a leave-one-out validation method, in which the training set consisted of all the trials of the data set except the one being tested. Note that analysing the different task epochs separately for the S-M electrodes produced classifier results that were also significantly above chance (auditory epoch, 40.2% ($\chi^2(1) = 47$, $P = 7 \times 10^{-12}$), production epoch, 23.2% ($\chi^2(1) = 5.6$, $P = 0.02$)).

To create the cumulative curves, the number of electrodes inputted into the classifier was increased linearly. To control for the variability in trial numbers, the minimum number of trials common to all subjects and electrodes was used. One-hundred iterations for each number of cumulative electrodes were performed, in which the specific trials and the specific electrodes were varied randomly and the number of SVD components was equal to the number of electrodes inputted to the classifier for the AUD and S-M electrodes, whereas five components were used for the PROD electrodes due to a lower number of components present in the PROD-electrode data.

Confusion matrix scores are simply the proportion of trials classified as the token on the horizontal axis (decoded) given that the actual trial came from the vertical axis (actual). Confusion matrices in Fig. 3a are shown for the largest number of cumulative electrodes in each electrode class.

To analyse the listen–speak transformation task responses (Fig. 4), the same decomposition (SVD) of the neural signal (70–160 Hz) was used. Note that instead of a seven-way classifier, a four-way classifier was used. Confusion matrices (Fig. 4C) are shown for the largest number of cumulative electrodes in each electrode class (AUD = 10; PROD = 19; S-M = 8). For the S-M electrodes, each response epoch (auditory, Fig. 4Cc; production, Fig. 4Cd) was analysed separately.

To measure the quality of each of the models (sensory, motor, sensory–motor or chance; Fig. 4d) we used the Kullback–Leibler divergence, which quantifies the amount of information lost in bits when Q (the model) is used to approximate P (the data):

$$D_{KL}(P||Q) = \sum_i P(i) \log_2 \left(\frac{P(i)}{Q(i)} \right)$$

where P is the classification percentage for each actual or decoded pair (see above) and Q is one of the four models: sensory, motor, sensory–motor and chance. The Kullback–Leibler divergence estimates the information distance between the pattern of classification errors predicted by each model, shown in Fig. 4B and the pattern of classification errors based on neural recordings, shown in Fig. 4C. Smaller Kullback–Leibler divergence reflects more information about classification errors and improved model fit. The sensory model (Fig. 4Ba) reflects classification scores that track the auditory speech input such that in both the match and the mismatch cases, the same input will be confused with each other. Conversely, the motor model (Fig. 4Bb) reflects classification scores that track the production output so that the same outputs will be confused with one another. However, the sensory–motor model (Fig. 4Bc) will reflect both the input and output such that classifications for each of the conditions presented (kig→kig, pob→pob, kig→pob and pob→kig) will be classified correctly. Finally, the chance model will simply reflect chance performance in all cases (0.25).

Classifier analysis of the cue data (‘Match Listen’ versus ‘Mismatch Listen’) in the listen–speak transformation task was analysed on the S-M electrodes for the subjects performing the task. The same linear classifier was used as above, but was performed during the cue period (0–1,000 ms post Cue) and was two-way (‘Match Listen’ versus ‘Mismatch Listen’ cues). The results demonstrated that the classification was not significant (mean classification = 52.3%, $\chi^2(1) = 0.08$, $P = 0.78$). Furthermore, using the same two-way classifier between the match and mismatch condition during the auditory epoch was also not significant (mean classification = 56.4%, $\chi^2(1) = 0.72$, $P = 0.4$). Taken together, this indicates that the sensory–motor transformations displayed by these electrodes cannot be due to a third population of neurons that code for the visual cue.

- Yang, A. I. *et al.* Localization of dense intracranial electrode arrays using magnetic resonance imaging. *NeuroImage* **63**, 157–165 (2012).
- Kovalev, D. *et al.* Rapid and fully automated visualization of subdural electrodes in the presurgical evaluation of epilepsy patients. *AJNR Am. J. Neuroradiol.* **26**, 1078–1083 (2005).
- Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I: segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
- Mitra, P. P. & Pesaran, B. Analysis of dynamic brain imaging data. *Biophys. J.* **76**, 691–708 (1999).
- Maris, E., Schoffelen, J.-M. & Fries, P. Nonparametric statistical testing of coherence differences. *J. Neurosci. Methods* **163**, 161–175 (2007).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Cell-autonomous correction of ring chromosomes in human induced pluripotent stem cells

Marina Bershteyn^{1,2*}, Yohei Hayashi^{3,4*}, Guillaume Desachy⁵, Edward C. Hsiao⁶, Salma Sami^{3,4}, Kathryn M. Tsang⁵, Lauren A. Weiss⁵, Arnold R. Kriegstein², Shinya Yamanaka^{3,4,7,8} & Anthony Wynshaw-Boris^{1,9}

Ring chromosomes are structural aberrations commonly associated with birth defects, mental disabilities and growth retardation^{1,2}. Rings form after fusion of the long and short arms of a chromosome, and are sometimes associated with large terminal deletions². Owing to the severity of these large aberrations that can affect multiple contiguous genes, no possible therapeutic strategies for ring chromosome disorders have been proposed. During cell division, ring chromosomes can exhibit unstable behaviour leading to continuous production of aneuploid progeny with low viability and high cellular death rate^{3–9}. The overall consequences of this chromosomal instability have been largely unexplored in experimental model systems. Here we generated human induced pluripotent stem cells (iPSCs)^{10–12} from patient fibroblasts containing ring chromosomes with large deletions and found that reprogrammed cells lost the abnormal chromosome and duplicated the wild-type homologue through the compensatory uniparental disomy (UPD) mechanism. The karyotypically normal iPSCs with isodisomy for the corrected chromosome outgrew co-existing aneuploid populations, enabling rapid and efficient isolation of patient-derived iPSCs devoid of the original chromosomal aberration. Our results suggest a fundamentally different function for cellular reprogramming as a means of ‘chromosome therapy’¹³ to reverse combined loss-of-function across many genes in cells with large-scale aberrations involving ring structures. In addition, our work provides an experimentally tractable human cellular system for studying mechanisms of chromosomal number control, which is of critical relevance to human development and disease.

We obtained fibroblasts from a Miller Dieker Syndrome (MDS) patient with ring chromosome 17, subsequently referred to as r(17). MDS is caused by heterozygous deletions of human band 17p13.3 (refs 14, 15) (Fig. 1a). This deletion alone leads to craniofacial dysmorphism, defective neuronal migration, abnormal cortical layering and nearly absent cortical folding with devastating neurological consequences such as mental retardation and intractable epilepsy^{14,16}. However, in this case the 17p13.3 deletion was in a ring chromosome, and the patient had a typical MDS phenotype¹⁴. To separate the effects of r(17) from the 17p13.3 deletion, we obtained fibroblasts from two additional MDS patients with similar deletions but without r(17) (Fig. 1b).

Two critical genes deleted in MDS are *PAFAH1B1* (encoding LIS1) and *YWHAE* (encoding 14-3-3 ϵ) (Fig. 1a)¹⁵. Consistent with this, MDS fibroblasts MDS1r(17), MDS2 and MDS3 (Fig. 1c) expressed reduced *PAFAH1B1* and *YWHAE* messenger RNA compared to control fibroblasts (Fig. 1d, e). MDS1r(17) fibroblasts had a 46,XY,r(17) karyotype in 95% of the cells (Figs 1i and 2a and Extended Data Fig. 1), with the remaining 5% exhibiting ring loss or secondary ring derivatives (Fig. 2b).

To investigate the behaviour of ring chromosomes in actively proliferating cells, we generated iPSCs using non-integrating episomal vectors¹⁷. All MDS iPSCs were morphologically indistinguishable from wild type (Fig. 1f) and expressed stem cell markers (Fig. 1g and Extended Data Fig. 2a–d). We confirmed that MDS iPSCs were free of exogenous factor integration (Extended Data Fig. 3a, b) and were functionally pluripotent, producing cell types of the three germ layers (Extended Data Figs 4 and 5).

We then analysed six early-passage MDS1r(17) iPSC clones for the presence of the ring (Fig. 1h) and, unexpectedly, found that four out of six clones did not have any detectable ring chromosomes (Fig. 1i), but grew well and had proper morphology (Extended Data Fig. 6a). In contrast, two other clones with detectable ring chromosomes (Fig. 1i) differentiated or stopped growing upon subsequent passaging (Extended Data Fig. 6b, c). Analysis of chromosome composition revealed that stable clones had 46 chromosomes and no ring in 85–100% of cells, in contrast to <15% of cells in unstable clones (Fig. 1i and Extended Data Fig. 6d, e). These results suggested that r(17) was incompatible with reprogramming and/or stem cell maintenance using our methods.

Further cytogenetic analysis of the first two MDS1r(17) iPSC clones demonstrated a normal 46,XY karyotype without r(17) (Fig. 2a–c). In addition, the deletion, which was readily detectable by Giemsa (G)-banding in MDS2 and MDS3 iPSCs, was not apparent in MDS1r(17) iPSCs (Fig. 2a and Extended Data Fig. 1). These findings could be explained by either clonal expansion of rare cells with a normal karyotype from mosaic fibroblasts; or repair or replacement of the ring chromosome during or after reprogramming. We reasoned that presence of a small fraction of cells with the deleted r(17) in predominantly corrected iPSC clones would confirm their origin from an abnormal fibroblast. To test this, we analysed interphase nuclei by fluorescence *in situ* hybridization (FISH) using a red probe to detect the sub-telomeric band 17p13.3 together with a green probe to detect the long-arm band 17q21.32 (Fig. 2d, e). Consistent with karyotype data, $\geq 80\%$ of cells in iPSC clones 1 and 2 demonstrated a normal signal pattern (2R2G) (Fig. 2f), suggesting that they have two intact copies of chromosome 17. However, 10–20% of interphase iPSCs had a 1R2G signal pattern, indicative of r(17) with p13.3 deletion (Fig. 2f). These results revealed persistent mosaicism in MDS1r(17) iPSCs, confirming their origin from ring-containing fibroblasts and supporting the hypothesis that repair or replacement of the ring chromosome occurred during reprogramming. Of note, despite several attempts (>120 cells analysed in multiple experiments), no evidence of r(17) was found in metaphase iPSCs, suggesting that in the pluripotent state such cells may be terminal and non-dividing. The expectation is that remaining interphase cells with r(17) would be gradually depleted.

¹Institute for Human Genetics and Department of Pediatrics, University of California, San Francisco, California 94143, USA. ²Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, California 94143, USA. ³Gladstone Institute of Cardiovascular Disease, San Francisco, California, 94158, USA. ⁴Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, California 94158, USA. ⁵Department of Psychiatry, Institute for Human Genetics, University of California, San Francisco, California 94143, USA. ⁶Division of Endocrinology and Metabolism and Institute for Human Genetics, Department of Medicine, University of California, San Francisco, California 94143, USA. ⁷Department of Anatomy, University of California, San Francisco, San Francisco, California 94143, USA. ⁸Department of Reprogramming Science, Center for iPS Cell Research and Application, Kyoto University, Kyoto 606-8507, Japan. ⁹Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA.

*These authors contributed equally to this work.

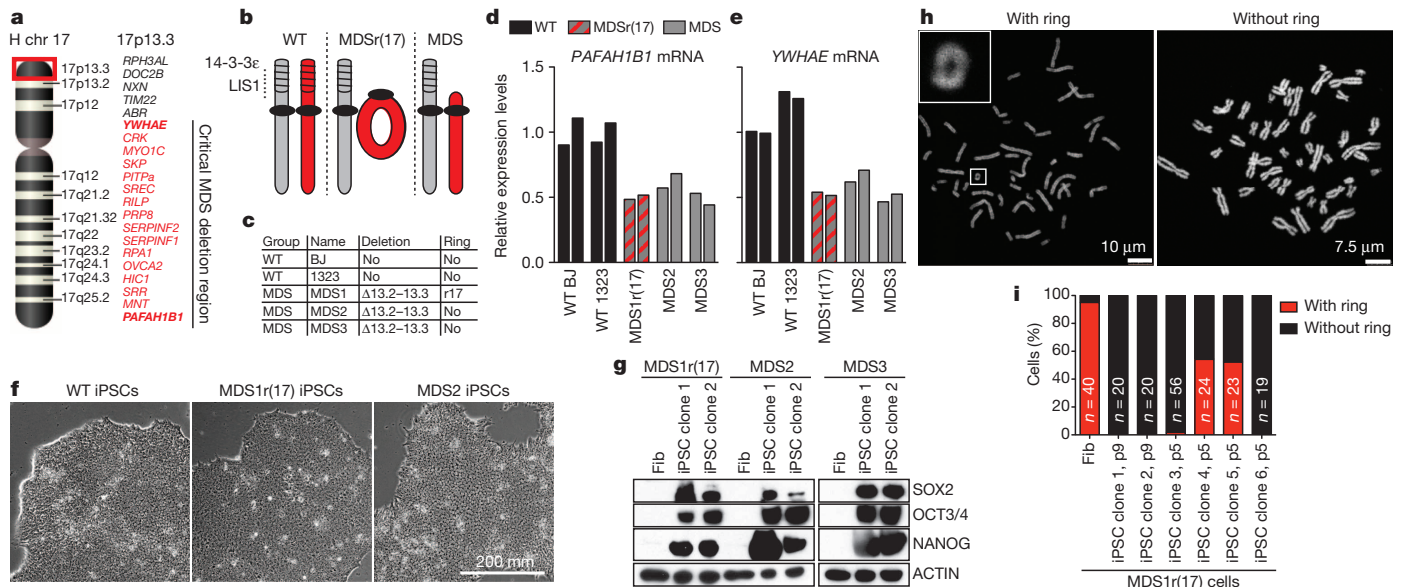


Figure 1 | Reprogramming from fibroblasts with r(17) produces multiple iPSC clones that do not have the ring chromosome. **a**, Schematic of human chromosome 17 (chr 17), highlighting band 17p13.3, which is deleted in MDS. Encoded genes are listed on the right, and genes within the minimal critical MDS deletion are shown in red. **b**, Schematic of chr 17 status in wild type (WT) and MDS cells. **c**, List of fibroblast lines used in this study. **d**, **e**, Quantitative

polymerase chain reaction with reverse transcription (RT-qPCR) for *PAFAH1B1* (**d**) and *YWHA6* (**e**) mRNA levels in fibroblasts ($n = 2$). **f**, Images of WT, MDS1r(17) and MDS2 iPSCs. **g**, Western blot analysis for pluripotency markers in fibroblasts (Fib) and two iPSC clones. **h**, Examples of metaphase spreads with or without the ring chromosome. **i**, Percentage of mitotic cells with or without the ring chromosome in fibroblasts or MDS1r(17) iPSC clones.

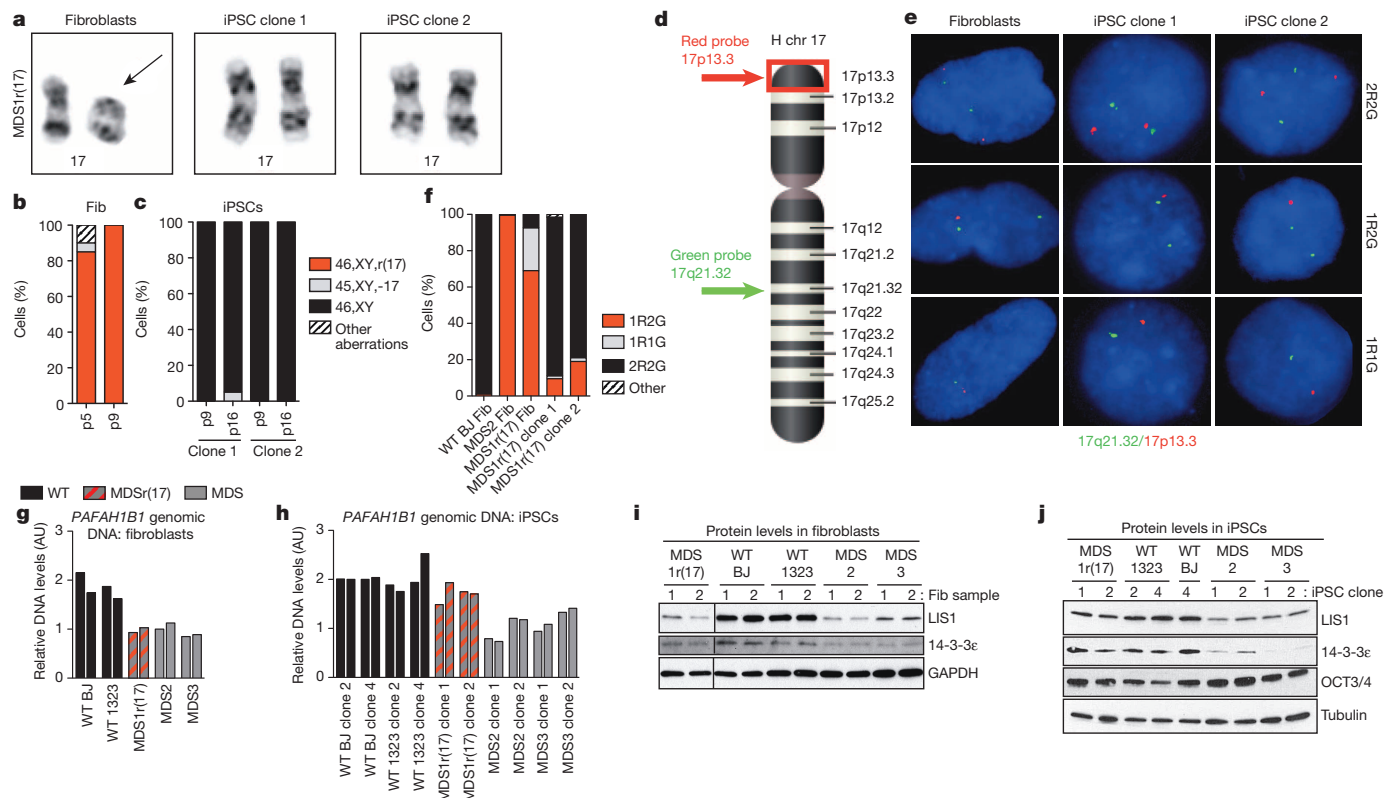


Figure 2 | Karyotypically normal cells predominate in early passage iPSC clones derived from MDS1r(17) fibroblasts. **a**, Images of chr 17 pairs in MDS1r(17) fibroblasts and iPSC clones. The arrow is pointing to r(17) only found in the fibroblasts. **b**, **c**, Proportion of MDS1r(17) fibroblasts (**b**) and iPSC clones 1 and 2 (**c**) with various karyotypes ($n = 20$ each). **d**, Approximate position of FISH probes on chr 17. **e**, Signal patterns obtained with FISH probes

in **d**, **f**, Proportions of cells with various signal patterns ($n = 200$ each). **g**, **h**, qPCR for genomic *PAFAH1B1* DNA in fibroblasts (**g**) and iPSCs (**h**). AU, arbitrary units; $n = 2$. **i**, **j**, Western blot analysis for LIS1 and 14-3-3e in fibroblasts (**i**) (two samples per fibroblast line collected on different days) and iPSCs (two clones per line) (**j**).

Cytogenetic and FISH results suggested that the MDS-associated deletion was rescued in MDS1r(17) iPSC clones 1 and 2. Consistent with that rescue, *PAFAH1B1* genomic DNA levels were restored from 50% of wild type in the original fibroblasts (Fig. 2g) to wild-type levels in iPSCs (Fig. 2h). Conversely, in MDS2 and MDS3 cells without r(17), *PAFAH1B1* genomic DNA levels were consistently reduced before and after reprogramming (Fig. 2g, h). Similar results demonstrating rescue were obtained by western blot analysis for LIS1 and 14-3-3 ϵ proteins (Fig. 2i, j). To investigate the status of the deleted locus before and after reprogramming in more detail, we performed single nucleotide polymorphism (SNP) microarrays. Analysis of total copy number for SNPs on chromosome 17 showed a clear decrease between positions 1 and 5,700,000 in MDS1r(17) fibroblasts, which was not observed in the corresponding iPSCs (Fig. 3a, b). In contrast, large terminal deletions were apparent in iPSCs derived from MDS2 and MDS3 fibroblasts (Fig. 3c, d). Of note, the extent of the deletion in MDS3 cells was exactly the same as in MDS1r(17) fibroblasts, making it an ideal control for the effect of the ring. Therefore, 17p13.3 deletion is recapitulated in iPSCs, and does not severely compromise pluripotent stem cell growth, viability or function. In contrast, r(17) appears to compromise *in vitro* stem cell maintenance and proliferation, and iPSCs that lost or repaired the ring chromosome showed preferential survival.

We proposed two possible mechanisms that could lead to loss of r(17) and rescue of the deleted locus. The first mechanism, called compensatory uniparental disomy (UPD), also referred to as monosomy correction, is known to occur during human development, and involves replacement of an abnormal chromosome with a copy of its normal

homologue¹⁸. This is a two-step process that could proceed through loss of r(17) due to anaphase lag, followed by duplication of the wild type chromosome 17 by nondisjunction (Fig. 3e). The second mechanism could occur through a ring opening owing to a double-strand break¹⁹, followed by homology-based repair or somatic recombination^{20,21} with the wild-type homologue (Fig. 3e). In both cases selection is likely to favour cells with 46 normal chromosomes. However, in the first scenario we would expect both chromosomes 17 in iPSCs to be isodisomic; that is, completely homozygous, whereas in the second scenario we would expect partial loss of heterozygosity (LOH) at the repaired locus with the rest of chromosome 17 being heterozygous (Fig. 3e). To distinguish between these possibilities, we used data from SNP arrays to determine heterozygosity on chromosome 17 in fibroblasts and iPSCs. Consistent with compensatory UPD, the two homologues of chromosome 17 were entirely homozygous in both iPSC clones derived from MDS1r(17) cells, whereas the original fibroblasts as well as MDS2 and MDS3 iPSCs exhibited 27% heterozygosity (Fig. 3f). This was a specific effect for the corrected chromosome 17 (Fig. 3g).

Non-random gain of a third chromosome 17 occurs frequently in long-term culture of human pluripotent cells owing to a proliferative growth advantage^{22–25}. To test whether chromosomes not typically gained or implicated in advantageous stem cell growth can replace the corresponding ring chromosome in iPSCs, we reprogrammed fibroblasts from two additional cases involving ring chromosome 13, subsequently referred to as r(13), with large deletions (Fig. 4a, f) associated with multiple congenital anomalies. We confirmed the presence of r(13) in approximately 80–100% of the fibroblasts, with the remaining cells exhibiting

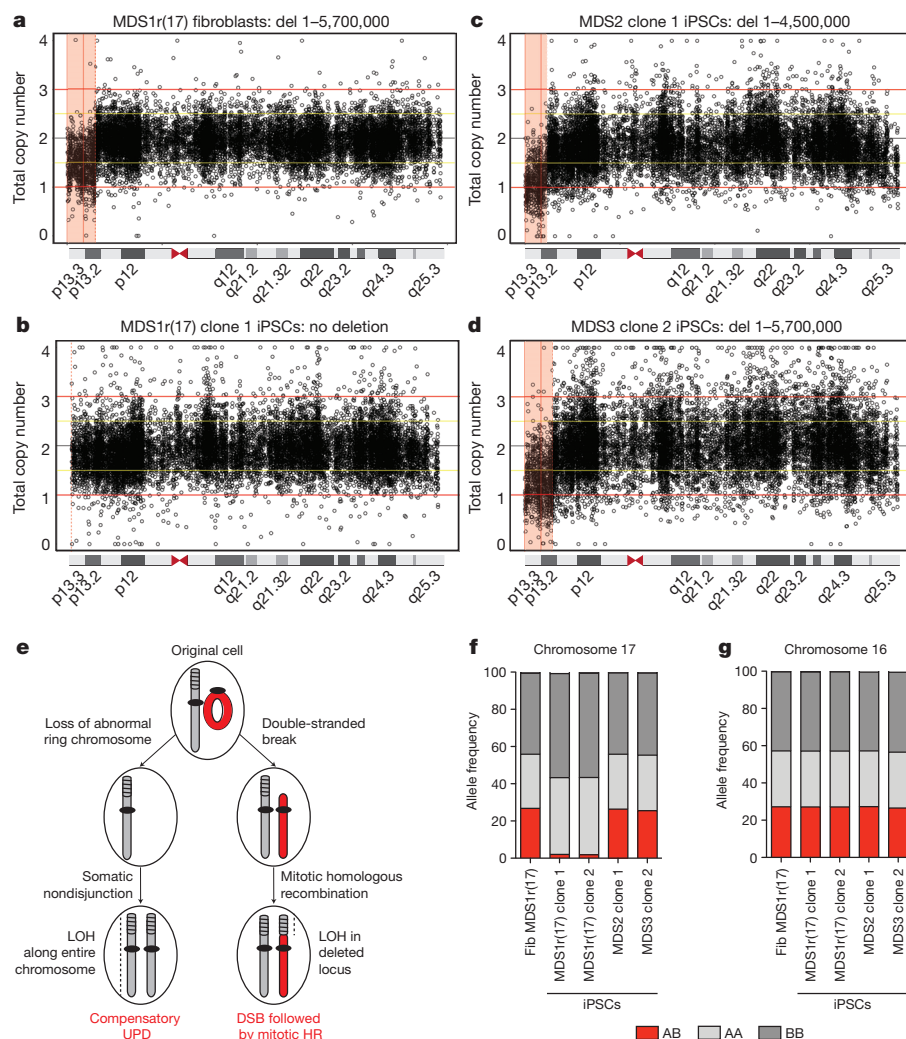


Figure 3 | Rescue of MDS-associated deletion in iPSCs derived from r(17) fibroblasts through compensatory uniparental isodisomy. a–d, Total copy number of SNPs across chr 17 in MDS1r(17) fibroblasts (a), MDS1r(17) iPSC clone 1 (b), MDS2 iPSC clone 1 (c), and MDS3 iPSC clone 2 (d). The pink shaded areas represent the deletions. e, Proposed mechanisms for how cells with r(17) end up with two intact chromosomes 17 after reprogramming. f, g, Frequency of heterozygous (red) or homozygous (shades of grey) SNPs on chr 17 (f) or chr 16 (g).

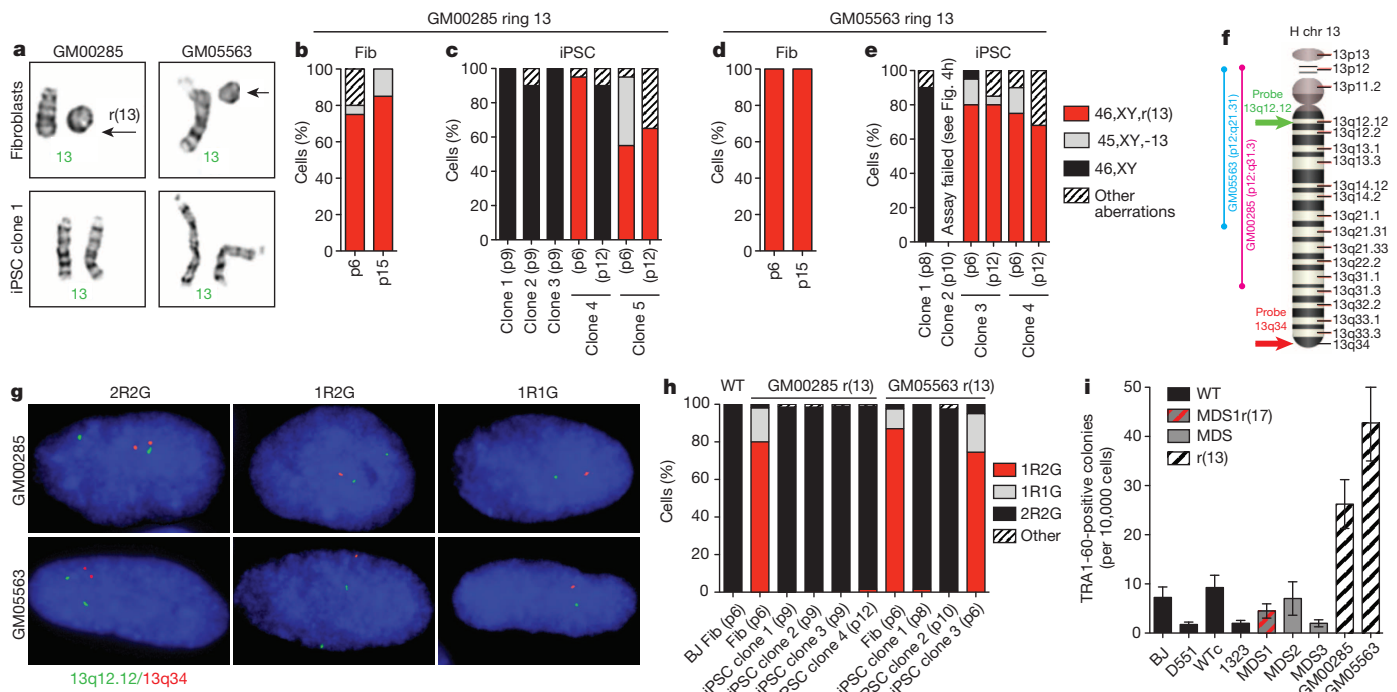


Figure 4 | Derivation of iPSC clones with a normal karyotype from fibroblasts with r(13). **a**, Images of chr 13 pairs in fibroblasts and repaired iPSCs from two individuals with r(13). The arrows are pointing to r(13) only found in the fibroblasts. **b–e**, Proportions of fibroblasts and iPSCs with various karyotypes ($n = 20$ each). **f**, Schematic of chr 13 showing approximate

45 chromosomes minus r(13) or non-clonal aberrations involving the ring structure (Fig. 4b, d and Extended Data Fig. 7a, d). Initial examination of iPSCs at passage six revealed 46,XY,r(13) karyotypes, although a fraction of cells had lost the ring chromosome (Fig. 4c, e). However, six out of nine clones that were examined after passage eight had a normal karyotype (Fig. 4a, c, e and Extended Data Fig. 7). In addition, GM00285 iPSC clone 4 transitioned from predominantly 46,XY,r(13) at passage six to 46,XY at passage twelve (Fig. 4c). These results demonstrate dynamic mosaicism between clonally related cells, leading to preferential survival of karyotypically normal iPSCs in a matter of six passages, which did not occur in the fibroblasts (Fig. 4b, d). In agreement, FISH studies revealed normal signal patterns in the corrected iPSC clones (Fig. 4f–h). Finally, SNP microarray analysis confirmed rescue of the deleted regions (Extended Data Fig. 8a–e) and showed that corrected iPSCs are completely homozygous for chromosome 13 (Extended Data Fig. 8f, g). The repaired iPSCs expressed stem cell markers (Extended Data Fig. 2e–j), differentiated into three germ layers (Extended Data Fig. 9), and were free of exogenous factor integration (Extended Data Fig. 3c, d). All of the established iPSC lines in this study were confirmed to be identical to their original fibroblasts by DNA fingerprinting (Extended Data Table 1).

Finally, to further rule out the possibility that repaired iPSC clones originated from rare 46,XY fibroblasts, we measured the reprogramming rates for the lines used in this study. If there is a bias against reprogramming from cells with ring chromosomes, then we would expect reduced reprogramming efficiency from ring-containing fibroblast populations. However, we found no significant difference between wild-type and MDS cells, while reprogramming efficiency from r(13) fibroblasts was increased compared to wild type (Fig. 4i). Therefore, r(17) or r(13) chromosomes do not interfere with episomal reprogramming. More likely, reprogramming increases cell divisions, increasing the probability that random nondisjunction compensates for the loss of the corresponding ring. Nondisjunction leading to chromosomal trisomy is the most commonly reported karyotypic change in cultured human ES cells and iPSCs²³, because mitotic checkpoints that ensure

proper alignment and disentanglement of sister chromatids before separation are less efficient than in differentiated cells^{23,26–28}. In this case, isodisomic cells have a growth advantage over cells with ring chromosomes including large deletions; however, they are not exactly the same as wild type. UPD could lead to undesirable clinical consequences through homozygosity for a recessive mutation or disruption of imprinting¹⁸. Thus, expression of imprinted genes and presence of recessive mutations would have to be carefully monitored and/or corrected for any therapeutic applications.

Compensatory UPD and dynamic mosaicism occur in actively proliferating cells *in vivo*, including cases involving ring chromosome 21, in which patients' lymphocytes transitioned from predominantly 46,XX,r(21) to 46,XX (with UPD21) several years later^{29,30}. These observations further suggest that taking advantage of this intrinsic property of iPSCs may be a feasible approach to correct combined loss-of-function and structural aberrations associated with additional ring chromosomes. Given that repair of large deletions involving multiple genes is not possible with current genome editing techniques, our results may have broad implications for regenerative medicine and advance the newly emerging concept of chromosome therapy¹³. More broadly, our findings underscore the importance of studying chromosome repair in stem cells and establish a novel platform for investigating the mechanisms that regulate chromosome number during early human development.

METHODS SUMMARY

The following primary human cells were used in this study: MDS1r(17) (GM06047), MDS2 (GM06097), MDS3 (GM09208), GM00285 and GM05563 (Coriell Institute for Medical Research); BJ and Detroit 551 (ATCC); HDF-1323 from a 48-year-old caucasian female (Cell Applications; 106-05a); HDF-WTc from a 30-year-old Japanese male, whose skin samples were collected at Gladstone. Human iPSCs were generated with the episomal Y4 plasmid mixture as described¹⁷ with some modifications (available in the online Methods). Established hiPSC lines were cultured in mTeSR1 medium (Stem Cell Technologies) on dishes coated with Matrigel (BD) or Synthemax (Corning). Chromosomal analyses were performed on 20 G-banded metaphase cells from each sample and FISH analyses were performed on two hundred interphase nuclei from each sample at Cell Line Genetics or the

University of California San Francisco cytogenetics laboratory. SNP genotyping to determine CNV presence and UPD status was performed using the Affymetrix Axiom EUR array platform by the Genomics Core Facility (GCF) at the University of California San Francisco for MDS samples and using the Agilent SurePrint G3 Human CGH + SNP 4 × 180K array platform by Cell Line Genetics for r(13) samples.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 September; accepted 29 November 2013.

Published online 12 January 2014.

- Cote, G. B. *et al.* The cytogenetic and clinical implications of a ring chromosome 2. *Ann. Genet.* **24**, 231–235 (1981).
- Kosztolányi, G. Does “ring syndrome” exist? An analysis of 207 case reports on patients with a ring autosome. *Hum. Genet.* **75**, 174–179 (1987).
- McClintock, B. The production of homozygous deficient tissues with mutant characteristics by means of the aberrant mitotic behavior of ring-shaped chromosomes. *Genetics* **23**, 315–376 (1938).
- Kistenmacher, M. L. & Punnett, H. H. Comparative behavior of ring chromosomes. *Am. J. Hum. Genet.* **22**, 304–318 (1970).
- Tommerup, N. & Lothe, R. Constitutional ring chromosomes and tumor suppressor genes. *J. Med. Genet.* **29**, 879–882 (1992).
- Jobanputra, V. *et al.* Changes in an inherited ring (22) due to meiotic recombination? Implications for genetic counseling. *Am. J. Med. Genet. A* **149A**, 1310–1314 (2009).
- Mantzouratau, A. *et al.* Meiotic and mitotic behaviour of a ring/deleted chromosome 22 in human embryos determined by preimplantation genetic diagnosis for a maternal carrier. *Mol. Cytogenet.* **2**, 3 (2009).
- Sodré, C. P. *et al.* Ring chromosome instability evaluation in six patients with autosomal rings. *Genet. Mol. Res.* **9**, 134–143 (2010).
- Zhang, H. Z. *et al.* Unique genomic structure and distinct mitotic behavior of ring chromosome 21 in two unrelated cases. *Cytogenet. Genome Res.* **136**, 180–187 (2012).
- Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).
- Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
- Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
- Jiang, J. *et al.* Translating dosage compensation to trisomy 21. *Nature* **500**, 296–300 (2013).
- Dobyns, W. B. *et al.* Miller-Dieker syndrome: lissencephaly and monosomy 17p. *J. Pediatr.* **102**, 552–558 (1983).
- Cardoso, C. *et al.* Refinement of a 400-kb critical region allows genotypic differentiation between isolated lissencephaly, Miller-Dieker syndrome, and other phenotypes secondary to deletions of 17p13.3. *Am. J. Hum. Genet.* **72**, 918–930 (2003).
- Wynshaw-Boris, A. *et al.* Lissencephaly: mechanistic insights from animal models and potential therapeutic strategies. *Semin. Cell Dev. Biol.* **21**, 823–830 (2010).
- Okita, K. *et al.* A more efficient method to generate integration-free human iPSC cells. *Nature Methods* **8**, 409–412 (2011).
- Robinson, W. P. Mechanisms leading to uniparental disomy and their clinical consequences. *Bioessays* **22**, 452–459 (2000).
- Speevak, M. D. *et al.* Molecular characterization of an inherited ring (19) demonstrating ring opening. *Am. J. Med. Genet. A* **121A**, 141–145 (2003).
- Hussein, S. M. *et al.* Genome damage in induced pluripotent stem cells: assessing the mechanisms and their consequences. *Bioessays* **35**, 152–162 (2013).
- Moynahan, M. E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature Rev. Mol. Cell Biol.* **11**, 196–207 (2010).
- Draper, J. S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nature Biotechnol.* **22**, 53–54 (2004).
- Baker, D. E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nature Biotechnol.* **25**, 207–215 (2007).
- Spits, C. *et al.* Recurrent chromosomal abnormalities in human embryonic stem cells. *Nature Biotechnol.* **26**, 1361–1363 (2008).
- Azuahata, T. *et al.* The inhibitor of apoptosis protein survivin is associated with high-risk behavior of neuroblastoma. *J. Pediatr. Surg.* **36**, 1785–1791 (2001).
- Damelin, M. *et al.* Decatenation checkpoint deficiency in stem and progenitor cells. *Cancer Cell* **8**, 479–484 (2005).
- Wilton, L. Preimplantation genetic diagnosis for aneuploidy screening in early human embryos: a review. *Prenat. Diagn.* **22**, 512–518 (2002).
- Vanneste, E. *et al.* Chromosome instability is common in human cleavage-stage embryos. *Nature Med.* **15**, 577–583 (2009).
- Petersen, M. B. *et al.* Uniparental isodisomy due to duplication of chromosome 21 occurring in somatic cells monosomic for chromosome 21. *Genomics* **13**, 269–274 (1992).
- Bartsch, O. *et al.* “Compensatory” uniparental disomy of chromosome 21 in two cases. *J. Med. Genet.* **31**, 534–540 (1994).

Acknowledgements The authors wish to thank N. Larocque and S. Fisher for access to the University of California San Francisco (UCSF) Human Embryonic Stem Cell Shared Research and Training Facility. We thank P. Tesar and Z. Nevin for comments on the manuscript; D. Srivastava and B. Conklin for advice; H. Belinson, A. Pollen and T. Nowakowski for helpful discussions; S. Hirotsune for LIS1 antibody; K. Essex for administrative support; and the Gladstone Stem Cell, Histology and Microscopy, and Bioinformatics Cores for technical support. The research was made possible by support from a NIH/NIGMS postdoctoral training grant in medical genetics (grant number GM007085-32) and a postdoctoral training fellowship from the California Institute for Regenerative Medicine (grant number TG2-01153) to M.B.; from the Uehara Memorial Foundation and USC’s Program for Breakthrough Biomedical Research to Y.H.; from the NIAMS/NIH (K08 AR056299) and the UCSF Department of Medicine to E.C.H.; and from the NHLBI/NIH (U01HL098179), the Leading Project of MEXT (Japan), the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) of the JSPS (Japan), Grants-in-Aid for Scientific Research of the JSPS and MEXT (Japan), the Program for Promotion of Fundamental Studies in Health Sciences of NIBIO (Japan), the L. K. Whittier Foundation, and the Roddenberry Foundation to S.Y. The Gladstone Institutes received support from a National Center for Research Resources Grant RR18928. S.Y. is a member without salary of the scientific advisory boards of iPierian, iPS Academia Japan, Megakaryon Corporation and HEALIOS K.K. Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of CIRM or any other agency of the State of California.

Author Contributions M.B. conceived and designed the study, generated and characterized MDS iPSC lines, performed experiments, analysed data, created the figures and wrote the manuscript. Y.H. helped to generate MDS iPSC lines, generated and characterized wild-type and r(13) iPSC lines, designed the study, performed experiments and analysed data. G.D. and L.A.W. performed SNP array genotyping and CNV calling analyses for MDS samples. E.C.H., S.S. and K.M.T. provided technical support for various experiments. A.R.K. provided advice, access to equipment and laboratory space for M.B. S.Y. supervised the study, provided advice, laboratory space and financial support. A.W.-B. supervised the study, provided advice, helped with design and interpretation, and provided laboratory space and financial support. A.W.-B., Y.H., S.Y., A.R.K., E.C.H. and L.A.W. edited the manuscript. All authors read and approved the final version of the manuscript.

Author Information The microarray data reported in this paper have been deposited to NCBI GEO with the accession numbers GSE52585 and GSE52691. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.W.-B. (ajw168@case.edu) or S.Y. (syamanaka@gladstone.ucsf.edu).

Citrullination regulates pluripotency and histone H1 binding to chromatin

Maria A. Christophorou^{1*}, Gonçalo Castelo-Branco^{1,2*}, Richard P. Halley-Stott^{1,3}, Clara Slade Oliveira^{1,4,5}, Remco Loos⁶, Aliaksandra Radziskeuskaya^{7,8}, Kerri A. Mowen⁹, Paul Bertone^{6,7,10}, José C. R. Silva^{7,8}, Magdalena Zernicka-Goetz^{1,5}, Michael L. Nielsen¹¹, John B. Gurdon^{1,3} & Tony Kouzarides^{1,12}

Citrullination is the post-translational conversion of an arginine residue within a protein to the non-coded amino acid citrulline¹. This modification leads to the loss of a positive charge and reduction in hydrogen-bonding ability. It is carried out by a small family of tissue-specific vertebrate enzymes called peptidylarginine deiminases (PADIs)² and is associated with the development of diverse pathological states such as autoimmunity, cancer, neurodegenerative disorders, prion diseases and thrombosis^{2,3}. Nevertheless, the physiological functions of citrullination remain ill-defined, although citrullination of core histones has been linked to transcriptional regulation and the DNA damage response^{4–8}. PADI4 (also called PAD4 or PADV), the only PADI with a nuclear localization signal⁹, was previously shown to act in myeloid cells where it mediates profound chromatin decondensation during the innate immune response to infection¹⁰. Here we show that the expression and enzymatic activity of Padi4 are also induced under conditions of ground-state pluripotency and during reprogramming in mouse. Padi4 is part of the pluripotency transcriptional network, binding to regulatory elements of key stem-cell genes and activating their expression. Its inhibition lowers the percentage of pluripotent cells in the early mouse embryo and significantly reduces reprogramming efficiency. Using an unbiased proteomic approach we identify linker histone H1 variants, which are involved in the generation of compact chromatin¹¹, as novel PADI4 substrates. Citrullination of a single arginine residue within the DNA-binding site of H1 results in its displacement from chromatin and global chromatin decondensation. Together, these results uncover a role for citrullination in the regulation of pluripotency and provide new mechanistic insights into how citrullination regulates chromatin compaction.

Pluripotent cells have the capacity to self-renew and differentiate into all somatic and germ-cell lineages and, hence, possess therapeutic potential for a multitude of medical conditions. Their generation by reprogramming of differentiated somatic cells has been achieved by nuclear transfer, cell fusion and transduction of transcription factors, such as Oct4 (also called Pou5f1), Sox2, Klf4 and c-Myc¹². Pluripotent cells have a distinctly open chromatin structure that is essential for unrestricted developmental potential^{13,14}, and reprogramming involves an almost complete epigenetic resetting of somatic cells¹³. The ability of PADI4-mediated histone citrullination to induce chromatin decondensation in neutrophils¹⁰ prompted us to ask whether it can have a role in pluripotency, where chromatin decondensation is also necessary. To investigate this we first assessed the expression of *Padi4* in the mouse embryonic stem-cell line ES Oct4-GIP (ES), the mouse neural

stem-cell line NSO4G and in induced pluripotent stem (iPS) cells derived from NSO4G (see Methods). *Padi4* is expressed in pluripotent ES and iPS cells but not multipotent neural stem cells (Fig. 1a). Culture of ES cells in 2i/LIF medium establishes a ground state of pluripotency¹⁵. This leads to the downregulation of lineage-specific markers and the upregulation of pluripotency factors, as well as rapid induction of *Padi4* expression (Fig. 1b). The pattern of *Padi4* expression follows closely that of *Nanog*, an essential transcription factor for the transition to ground-state pluripotency¹⁶ (Fig. 1a, b). Whereas other *Padi* genes are expressed in pluripotent cells, *Padi4* is the only one for which expression clearly associates with naive pluripotency (Extended Data Fig. 1a, b). Citrullination of histone H3 (H3Cit), a modification shown previously to be carried out specifically by PADI4 (ref. 17), is detectable in ES and iPS cells (Extended Data Fig. 1c), indicating that Padi4 is also enzymatically active. H3 and global citrullination are undetectable in NSO4G cells (Extended Data Fig. 1c, d).

To determine the kinetics of Padi4 activation during the establishment of pluripotency, we examined RNA and protein samples collected daily during the course of reprogramming of NSO4G into iPS cells¹⁶. *Padi4* is rapidly induced in NSO4G after transduction of reprogramming factors but only becomes active to citrullinate H3 after introduction of 2i/LIF, closely following the onset of *Nanog* expression (Fig. 1c). These observations strongly suggested that Padi4 activity is associated with ground-state pluripotency and prompted us to examine whether *Padi4* is part of the pluripotency transcriptional network.

First, we asked whether the reprogramming factors regulate *Padi4* expression, using the ZHBTc4.1 and 2T522C cell lines where Oct4 and Sox2, respectively, can be deleted acutely in response to doxycycline treatment (see Methods). Deletion of Oct4, but not Sox2, led to a decrease in *Padi4* messenger RNA levels (Extended Data Fig. 1e). Furthermore, whereas Oct4 and Klf4 occupy the *Padi4* promoter in ES, but not NSO4G, cells, Sox2 is bound in both cell types (Extended Data Fig. 1f). To understand the effects of Padi4 on transcriptional regulation in pluripotent cells, we analysed the transcriptome of ES cells upon overexpression of human PADI4 and inhibition of endogenous Padi4. Several key pluripotency genes are upregulated in response to PADI4 overexpression (Fig. 1d, Extended Data Fig. 2a and Supplementary Table 2), including *Klf2*, *Tcl1*, *Tcfap2c* and *Kit*. *Tcl1* was previously identified as the only regulator of self-renewal upregulated in ground-state pluripotency¹⁵, and overexpression of *Tcl1* or *Tcfap2c* positively influences reprogramming¹⁸. Gene Ontology (GO) analysis of this data set indicates an enrichment of genes involved in stem-cell development and maintenance (Fig. 1e). In addition, knockdown

¹The Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. ²Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden. ³Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. ⁴EMBRAPA Dairy Cattle Research Center, Juiz de Fora, Brazil.

⁵Department of Physiology, Development and Neuroscience, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. ⁶European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. ⁷Wellcome Trust—Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. ⁸Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. ⁹Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA. ¹⁰Genome Biology and Developmental Biology Units, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany. ¹¹Department of proteomics, The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Faculty of Health Sciences, Blegdamsvej 3b, DK-2200 Copenhagen, Denmark. ¹²Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK.

*These authors contributed equally to this work.

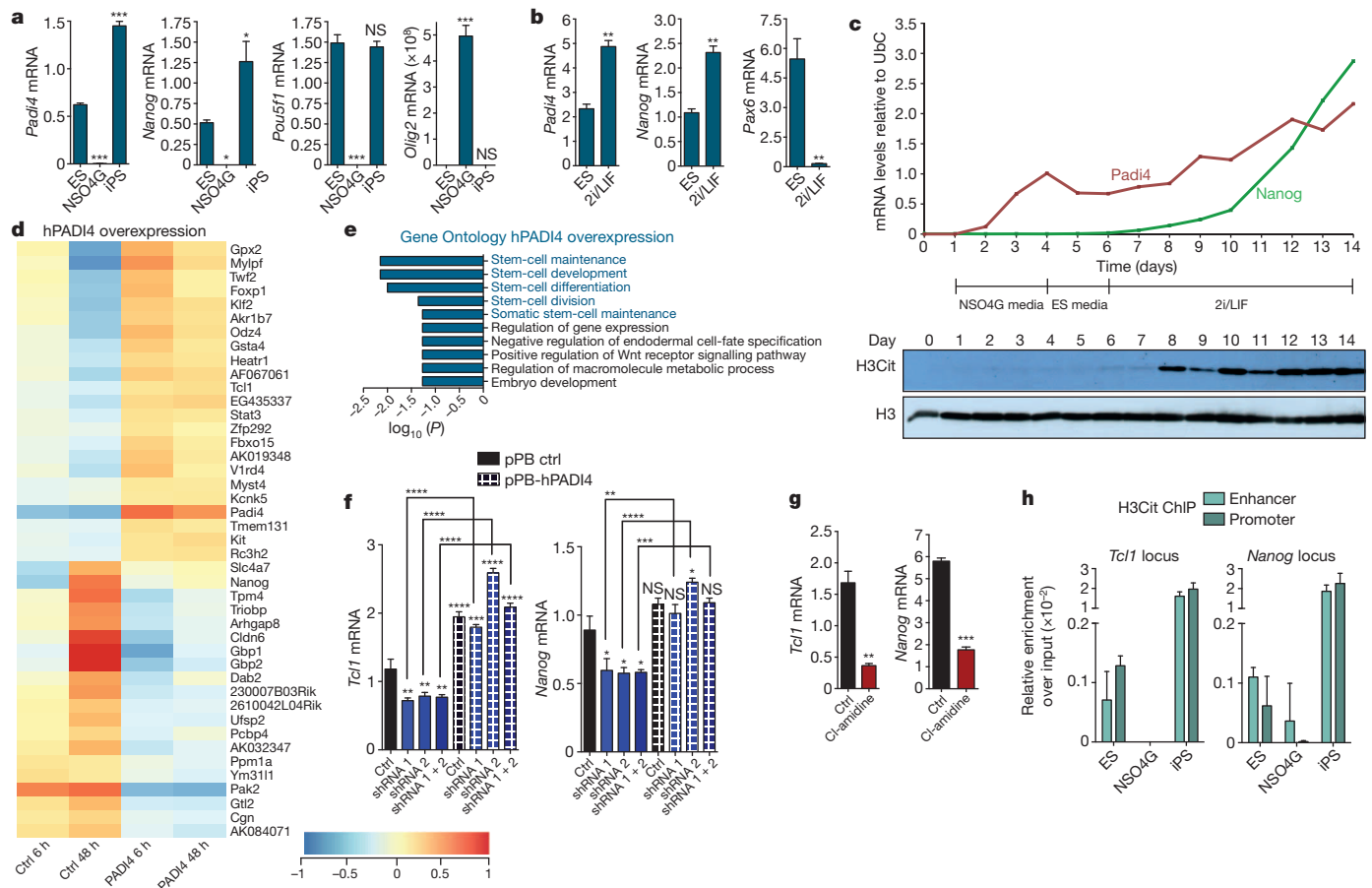


Figure 1 | Padi4 expression and activity are features of pluripotent cells.

a, **b**, qRT-PCR for *Padi4* and *Nanog* expression in ES, NSO4G and iPS cells (a), and in ES cells on culture in 2i/LIF for one passage (b). *Pou5f1*, *Olig2* and *Pax6* are presented as controls. Expression normalized to ubiquitin (*Ubc*). Error bars indicate standard error of the mean of three biological replicates.

c, qRT-PCR for *Padi4* and *Nanog* expression and H3Cit immunoblot during the course of reprogramming (see also Extended Data Fig. 5a). Loading control: total histone H3. Representative of four experiments.

d, Heat map of the genes regulated on human PADI4 (hPADI4) overexpression in mouse ES cells, as determined by microarray analysis. Displayed values are normalized log intensities, minus the mean expression of the gene across the four samples. Hierarchical clustering based on correlation.

e, Gene Ontology for Biological Process (GOBP) analysis of the above microarray data set. *P* value is corrected

of *Padi4* in mouse ES cells leads to decreased expression of *Tcf1* and *Nanog*, which is rescued by exogenous expression of RNA interference (RNAi)-resistant human PADI4 (Fig. 1f and Extended Data Fig. 2b, c). These genes are under the control of PADI4 enzymatic activity, as treatment with the chemical inhibitor Cl-amidine, which disrupts citrullination by PADI4 (ref. 19), downregulates their expression (Fig. 1g). Chromatin immunoprecipitation (ChIP) analysis indicated that H3Cit is present on regulatory regions of *Tcf1* and *Nanog* in ES and iPS cells, but not NSO4G (Fig. 1h and Extended Data Fig. 3a). Accordingly, exogenously expressed human PADI4 localizes to and is enzymatically active on these regions, as well as regulatory regions of *Klf2* and *Kit* in ES cells (Extended Data Fig. 3b, c). In contrast to PADI4 overexpression, treatment of ES cells with Cl-amidine led to upregulation of differentiation markers such as *Prickle1*, *Epha1* and *Wnt8a* and downregulation of pluripotency markers such as *Klf5* (Extended Data Fig. 4a, b and Supplementary Table 3), in addition to *Nanog* and *Tcf1* (Fig. 1g). GO analysis of this data set indicated enrichment in genes involved in cell differentiation (Extended Data Fig. 4c). *Pou5f1*, *Klf4*, *Sox2* and *c-Myc* were not affected by PADI4 modulation (Extended Data Fig. 2a and Supplementary Tables 2 and 3). Cumulatively, the above results place Padi4 within the pluripotency

transcriptional network, indicating that it acts downstream of some of the cardinal reprogramming factors to regulate a specific subset of pluripotency genes.

Prompted by the above observations, we investigated whether Padi4 is necessary for pluripotency, as assessed during reprogramming (Extended Data Fig. 5a) and in the early stages of embryo development. NSO4G cells express a green fluorescent protein (GFP) reporter under the control of the Oct4 regulatory sequences, which is activated on acquisition of pluripotency¹⁶, allowing us to trace reprogrammed cells. Knockdown of *Padi4* in NSO4G-derived pre-iPS cells impaired the ability of the cells to establish H3Cit upon switch to 2i/LIF medium and led to a clear reduction in reprogramming (Fig. 2a, Extended Data Fig. 5b–e and Supplementary Video 1). Consistent with this finding, levels of *Tcf1* and *Nanog* were not elevated on reprogramming to the same extent as in control cells (Fig. 2b and Extended Data Fig. 5f). Cl-amidine treatment led to a marked reduction of reprogramming efficiency and H3Cit (Fig. 2c and Extended Data Fig. 5g–i), indicating that the catalytic activity of Padi4 is important for the induction of pluripotency.

Padi4 expression and H3Cit are detected in the early embryo^{20,21} and *Padi4*-null mice are born in lower numbers than would be expected by Mendelian inheritance²², indicating that Padi4 loss affects embryonic

transcriptional network, indicating that it acts downstream of some of the cardinal reprogramming factors to regulate a specific subset of pluripotency genes.

Prompted by the above observations, we investigated whether Padi4 is necessary for pluripotency, as assessed during reprogramming (Extended Data Fig. 5a) and in the early stages of embryo development. NSO4G cells express a green fluorescent protein (GFP) reporter under the control of the Oct4 regulatory sequences, which is activated on acquisition of pluripotency¹⁶, allowing us to trace reprogrammed cells. Knockdown of *Padi4* in NSO4G-derived pre-iPS cells impaired the ability of the cells to establish H3Cit upon switch to 2i/LIF medium and led to a clear reduction in reprogramming (Fig. 2a, Extended Data Fig. 5b–e and Supplementary Video 1). Consistent with this finding, levels of *Tcf1* and *Nanog* were not elevated on reprogramming to the same extent as in control cells (Fig. 2b and Extended Data Fig. 5f). Cl-amidine treatment led to a marked reduction of reprogramming efficiency and H3Cit (Fig. 2c and Extended Data Fig. 5g–i), indicating that the catalytic activity of Padi4 is important for the induction of pluripotency.

Padi4 expression and H3Cit are detected in the early embryo^{20,21} and *Padi4*-null mice are born in lower numbers than would be expected by Mendelian inheritance²², indicating that Padi4 loss affects embryonic

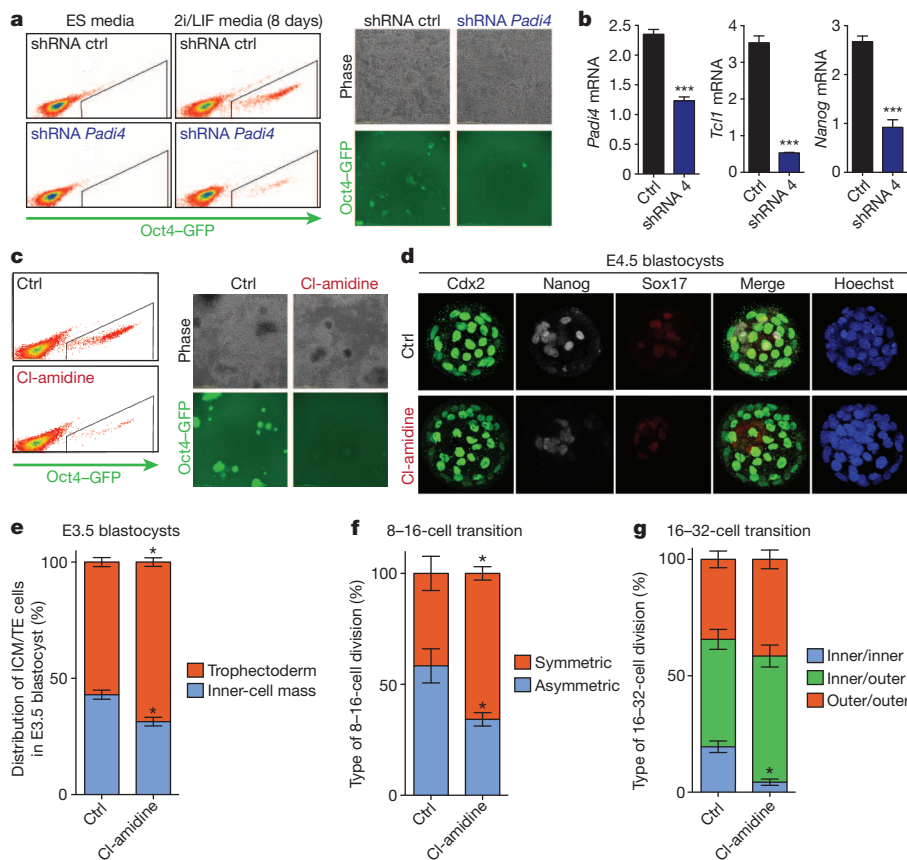


Figure 2 | Citrullination and Padi4 regulate pluripotency during reprogramming and early embryo development. **a**, Flow cytometry analysis and phase contrast/fluorescence images for the assessment of Oct4–GFP reporter expression after reprogramming of pre-iPS cells stably expressing *Padi4* and control shRNAs. Data are representative of four independent experiments. See Supplementary Video 1 for time-lapse video. **b**, qRT-PCR for expression of *Tcl1*, *Nanog* and *Padi4* at the end of the above reprogramming assay. Error bars indicate standard error of the mean of triplicate samples. **c**, Flow cytometry analysis and phase contrast/fluorescence images for the assessment of Oct4–GFP reporter expression after reprogramming assay of pre-iPS cells treated with 200 μ M Cl-amidine. Representative of three independent experiments. **d**, E4.5 blastocysts from 2-cell-stage embryos treated with 10 μ M Cl-amidine. Sox17 (primitive endoderm marker, red), Cdx2 (trophoblast marker, green), Nanog (epiblast marker, white) and Hoechst 33342 (blue) images are shown. **e**, Distribution of inner-cell mass (ICM) versus trophoblast (TE) cells in E3.5 blastocyst treated as above. **f**, **g**, Time-lapse analysis of embryos in 10 μ M Cl-amidine from 2-cell stage. Number of symmetric versus asymmetric divisions at the 8–16-cell transition (**f**) and type of divisions at the 16–32-cell transition (**g**) is shown. Error bars indicate standard error of the mean. Statistical significance was determined by unpaired *t*-test (**b**), or Mann–Whitney *U*-test on non-normal distribution (**e–g**). Asterisks denote difference with control; **P* \leq 0.05, ***P* \leq 0.01, ****P* \leq 0.001, *****P* \leq 0.0001.

development. To assess the role of Padi4 in early development, we cultured mouse embryos in Cl-amidine-containing medium from the 2-cell stage and throughout pre-implantation development (see Methods and Extended Data Fig. 6a–c). Using 200 μ M Cl-amidine resulted in a complete developmental arrest of the embryos at the 8-cell stage (Extended Data Fig. 6a). We therefore used the maximum dose of Cl-amidine that reduced H3Cit (Extended Data Fig. 6b, c) but did not induce arrest (10 μ M). This led to a reduced percentage of pluripotent Nanog-positive epiblast cells and an increased percentage of differentiated trophoblast cells at the blastocyst stage (Fig. 2d, e and Extended Data Fig. 6d, e). Time-course analyses of the cleavage patterns and cell-fate decisions in early embryos showed that Cl-amidine increased the number of symmetric cell divisions at the expense of asymmetrical divisions at the 8–16- and 16–32-cell transitions (Fig. 2f, g). This resulted in 16-cell-stage embryos with fewer inner cells (destined for pluripotency) and greater numbers of outer cells (destined for differentiation into trophoblast; reviewed in ref. 23) (Extended Data Fig. 6f). Treatment with another Padi4 inhibitor, Thr-Asp-F-amidine (TDFa)²⁴, but not the HDAC inhibitor trichostatin A (TSA), had similar effects (Extended Data Figs 7 and 8). These results indicate that Padi4 activity also promotes the maintenance of pluripotent cells in the early mouse embryo.

To elucidate the molecular mechanisms by which Padi4 regulates pluripotency, we aimed to identify PADI4 substrates in the chromatin fraction of mouse ES cells using stable isotope labelling of amino acids in cell culture (SILAC) (Fig. 3a and Extended Data Fig. 9a, b). Among the identified PADI4 substrates were Atrx, Dnmt3b, Trim28 and variants of linker histone H1 (Fig. 3b–e, Extended Data Figs 9c–f and 10a, b, and Supplementary Table 4), all of which can have an impact on pluripotency. Histone H1 stabilizes the nucleosome and facilitates chromatin condensation, a state that is less permissive to processes that require access to the DNA, such as transcription¹¹. The identified citrullinated H1 peptides correspond to, and are common between, variants

H1.2, H1.3 and H1.4 (Fig. 3d, e), whereas an additional peptide corresponds to the same residue in H1.5 (Extended Data Fig. 10a, b). In ES cells, H1.2, H1.3 and H1.4 are required for chromatin compaction²⁵, whereas their depletion leads to increased expression of pluripotency genes such as *Nanog* and stalls them in a self-renewal state with impaired differentiation capability²⁶. Notably, H1 is more loosely bound to chromatin in ES cells than in differentiated cells¹⁴ and its genomic localization in cancer cells was shown to anti-correlate with that of PADI4 (ref. 7). Mass spectrometric analysis accounted for all arginine residues within H1 but indicated that Arg 54 is the only site citrullinated by PADI4 (Fig. 3e and Extended Data Fig. 9c). Indeed, we found that although H1.2 is citrullinated in ES cells (by endogenous Padi4, and significantly increased on PADI4 overexpression), it is refractory to modification when Arg 54 is mutated (Fig. 3f). Similar results were obtained in *in vitro* citrullination assays (Extended Data Fig. 10c). H1 Arg 54 lies within the globular domain of H1 (Extended Data Fig. 10d), which is highly conserved among the linker histone family and is necessary for interaction with nucleosomal DNA^{27,28}. To test whether H1 Arg 54 is necessary for binding of H1 to nucleosomes, we mutated and assessed it in nucleosome-binding assays. Figure 3g shows that an R54A mutant, which mimics the charge change that accompanies citrullination, is impaired for nucleosome binding. An R54K mutant, which retains the positive charge, is impaired to a lesser extent (Fig. 3g), indicating that H1R54 is important for electrostatic interactions between H1.2 and the nucleosome. The above results open up the possibility that PADI4 may affect chromatin compaction in pluripotent cells. To test this hypothesis, we first assessed whether citrullination by ectopic PADI4 can lead to decondensation of differentiated cell chromatin. Recombinant PADI4 was added to permeabilized and stabilized differentiated C2C12 mouse myoblast nuclei (Fig. 4a). This protocol ensures stabilization of the nuclear component while allowing the free diffusion of non-chromatin-bound nuclear proteins into the extranuclear fraction, and their collection by washing. Incubation with active PADI4

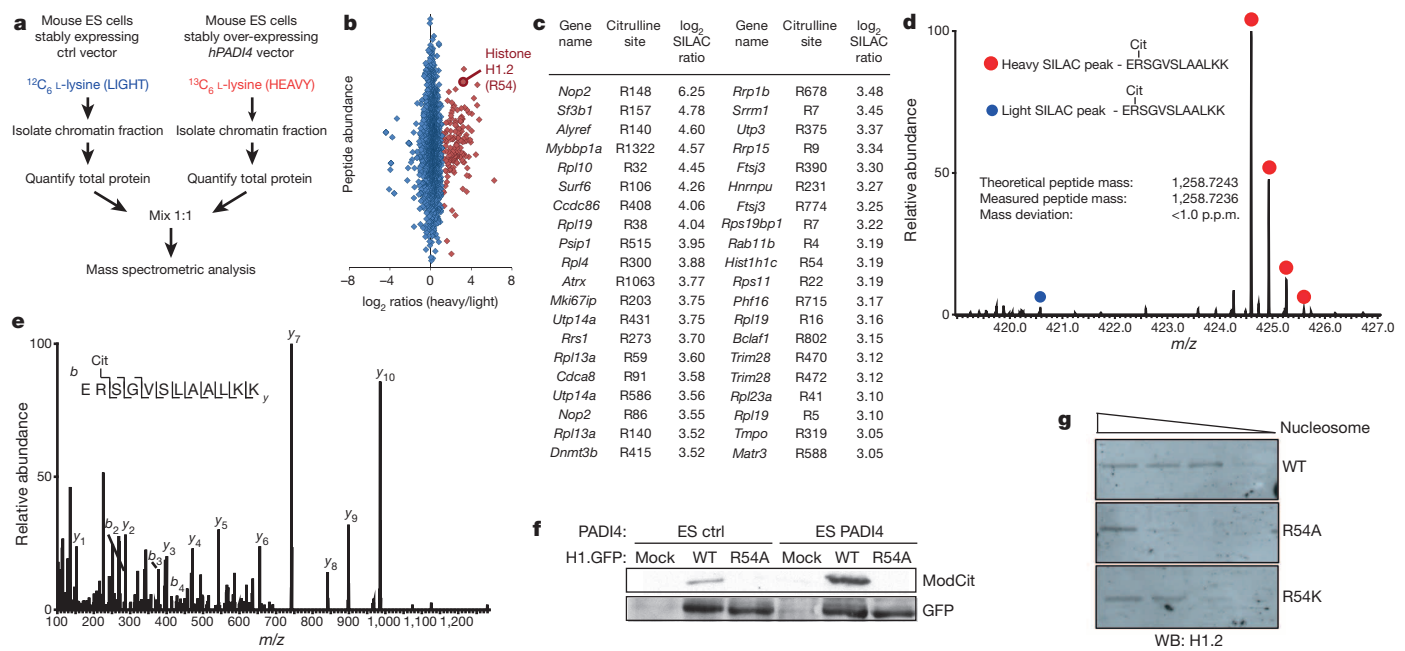


Figure 3 | PADI4 citrullinates Arg 54 on linker histone H1 and affects its binding to nucleosomal DNA. **a**, Experimental strategy for screening for PADI4 citrullination substrates in the chromatin fraction of ES cells. **b**, Scatter plot representing the overall fold change for all identified citrullination sites. Red diamonds indicate PADI4-regulated citrullinations. **c**, Table representing the 40 most highly regulated PADI4 substrates, their individual citrullination sites and the log₂ SILAC ratio. Complete data set is available in Supplementary Table 4. **d**, Quantification of citrullination site Arg 54 on H1.2 through

differentiation regulation of the triply charged peptide ERSVSLAALKK. **e**, Fragmentation spectra of the triply charged and heavy SILAC labeled LysC peptide ERSVSLAALKK surrounding Arg 54 of H1.2. The y and b series indicate fragments at amide bonds of the peptide. **f**, Citrullination immunoblot of wild-type (WT) and R54A mutant GFP-tagged H1.2 expressed and pulled down from ES cells expressing PADI4 or control vector (mock). Control for the efficiency of the pull down: GFP. **g**, Nucleosome pull-down assay using wild-type and Arg 54-mutant H1.2. WB, western blot.

(Extended Data Fig. 10e, f) leads to the eviction of H1 from the chromatin and its diffusion out of the permeabilized nucleus (Fig. 4b). The evicted H1 is citrullinated on Arg 54, as determined by mass spectrometry (Extended Data Fig. 10g, h). Consistent with this, PADI4-treated cells showed evidence of decondensed chromatin, as determined by nuclear swelling, diffuse 4',6-diamidino-2-phenylindole (DAPI) staining and increased sensitivity to micrococcal nuclease (Fig. 4c, d and Extended Data Fig. 10i). Similar results were observed when PADI4 was overexpressed in C2C12 cells (Fig. 4e) or NSO4G cells (data not shown). To monitor whether Padi4 can affect H1 binding on pluripotent

cell chromatin, we performed ChIP-qPCR (quantitative PCR) analyses of H1.2 on the regulatory regions of *Tcl1* and *Nanog* and found that it is stabilized on *Padi4* knockdown (Fig. 4f). The ability of Padi4 to disrupt the binding of H1 to nucleosomal DNA provides a novel mechanistic example of how citrullination regulates protein function and chromatin condensation.

The work described here identifies citrullination of chromatin components by Padi4 as a feature of pluripotency (Fig. 4g), in addition to its previously described role in the myeloid lineage. One of the reasons for the restricted expression pattern of Padi4 may be the requirement for

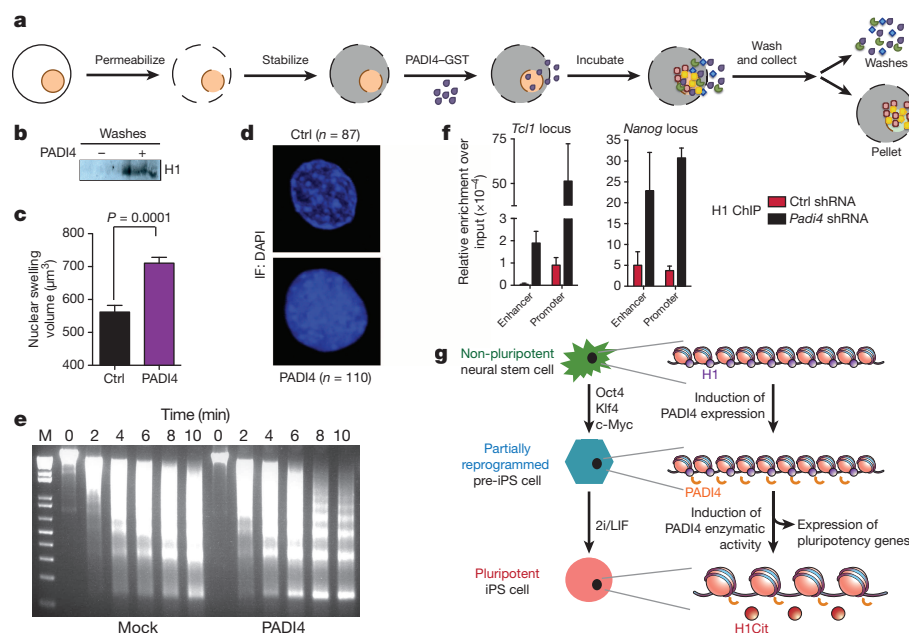


Figure 4 | PADI4 evicts histone H1 from chromatin and affects chromatin condensation. **a**, Schematic representation of treatment of C2C12 myoblast nuclei with recombinant PADI4.

b, Immunoblot analysis of the wash fraction after the above treatment, for histone H1.2. **c**, **d**, Quantification of nuclear volume (**c**) and representative DAPI fluorescence (**d**) on treatment of permeabilized C2C12 nuclei with recombinant PADI4. Error bars indicate standard error of the mean. Statistical significance determined by unpaired student's *t*-test. **e**, Micrococcal nuclease digestion of C2C12 cells overexpressing an empty vector (mock) or human PADI4. M, size marker. **f**, ChIP-qPCR for H1.2 on the regulatory regions of *Tcl1* and *Nanog* in mouse ES cells stably expressing *Padi4* or control shRNA. Error bars indicate standard error of the mean of three technical qPCR replicates. Representative of two experiments. **g**, Proposed model for the role of PADI4 in the regulation of pluripotency.

an open chromatin state in these cell types. The selective expression characteristics and the inducible nature of the catalytic activity of Padi4 indicate that it is under tight spatial and temporal regulation, giving it a unique status among chromatin-modifying enzymes. As such, inappropriate Padi4 activity may have deleterious consequences, which may explain its activation in cancers of varying origin during progression to malignancy²⁹. Indeed, citrullination is a common feature of several unrelated diseases, suggesting that strict regulation is probably a requirement for the physiological function of all PADIs. During review of this manuscript, a study suggested that PADI2, thought to be mainly cytoplasmic, can also citrullinate histones and lead to transcriptional activation³⁰. This opens up the possibility that other PADIs may mediate nuclear events in specific contexts, including in pluripotent cells. Further research into the function and targets of PADIs is likely to illuminate the aetiology of several pathologies.

METHODS SUMMARY

Information regarding cell lines, antibodies, plasmids and chemical inhibitors used in this study, as well as detailed protocols for reprogramming and cell culture, overexpression and knockdown, gene expression analysis (qRT-PCR and microarray), chromatin immunoprecipitation, embryo collection, culture, immunofluorescence staining and analysis, immunoblot analysis, SILAC labelling, mass spectrometry and identification of citrullinated peptides and proteins, native purification of wild-type and mutant linker histone H1.2, H1-GFP and nucleosome pull-down assays, purification of recombinant PADI4-GST and *in vitro* deimination assays, treatment of permeabilized cells with recombinant PADI4, and micrococcal nuclease digestion are provided in the Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 September 2012; accepted 6 December 2013.

Published online 26 January 2014.

- Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. & Pruijn, G. J. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *BioEssays* **25**, 1106–1118 (2003).
- Wang, S. & Wang, Y. Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis. *Biochim. Biophys. Acta* **1829**, 1126–1135 (2013).
- Martinod, K. *et al.* Neutrophil histone modification by peptidylarginine deiminase 4 is critical for deep vein thrombosis in mice. *Proc. Natl Acad. Sci. USA* **110**, 8674–8679 (2013).
- Hagiwara, T., Nakashima, K., Hirano, H., Senshu, T. & Yamada, M. Deimination of arginine residues in nucleophosmin/B23 and histones in HL-60 granulocytes. *Biochem. Biophys. Res. Commun.* **290**, 979–983 (2002).
- Cuthbert, G. L. *et al.* Histone deimination antagonizes arginine methylation. *Cell* **118**, 545–553 (2004).
- Wang, Y. *et al.* Human PAD4 regulates histone arginine methylation levels via demethylation. *Science* **306**, 279–283 (2004).
- Zhang, X. *et al.* Genome-wide analysis reveals PADI4 cooperates with Elk-1 to activate c-Fos expression in breast cancer cells. *PLoS Genet.* **7**, e1002112 (2011).
- Tanikawa, C. *et al.* Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. *Nature Commun.* **3**, 676 (2012).
- Asaga, H., Nakashima, K., Senshu, T., Ishigami, A. & Yamada, M. Immunocytochemical localization of peptidylarginine deiminase in human eosinophils and neutrophils. *J. Leukoc. Biol.* **70**, 46–51 (2001).
- Neeli, I., Khan, S. N. & Radic, M. Histone deimination as a response to inflammatory stimuli in neutrophils. *J. Immunol.* **180**, 1895–1902 (2008).
- Buttinelli, M., Panetta, G., Rhodes, D. & Travers, A. The role of histone H1 in chromatin condensation and transcriptional repression. *Genetica* **106**, 117–124 (1999).
- Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
- Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nature Rev. Mol. Cell Biol.* **12**, 36–47 (2011).
- Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
- Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
- Theunissen, T. W. *et al.* Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr. Biol.* **21**, 65–71 (2011).
- Darrah, E., Rosen, A., Giles, J. T. & Andrade, F. Peptidylarginine deiminase 2, 3 and 4 have distinct specificities against cellular substrates: novel insights into autoantigen selection in rheumatoid arthritis. *Ann. Rheum. Dis.* **71**, 92–98 (2012).
- Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
- Luo, Y. *et al.* Inhibitors and inactivators of protein arginine deiminase 4: functional and structural characterization. *Biochemistry* **45**, 11727–11736 (2006).
- Brahmajosyula, M. & Miyake, M. Localization and expression of peptidylarginine deiminase 4 (PAD4) in mammalian oocytes and preimplantation embryos. *Zygote* **21**, 314–324 (2013).
- Kan, R. *et al.* Potential role for PADI-mediated histone citrullination in preimplantation development. *BMC Dev. Biol.* **12**, 19 (2012).
- Li, P. *et al.* PAD4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps. *J. Exp. Med.* **207**, 1853–1862 (2010).
- Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nature Rev. Genet.* **10**, 467–477 (2009).
- Jones, J. E. *et al.* Synthesis and screening of a haloacetamide containing library to identify PAD4 selective inhibitors. *ACS Chem. Biol.* **7**, 160–165 (2012).
- Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**, 1199–1212 (2005).
- Zhang, Y. *et al.* Histone H1 depletion impairs embryonic stem cell differentiation. *PLoS Genet.* **8**, e1002691 (2012).
- Goytisolo, F. A. *et al.* Identification of two DNA-binding sites on the globular domain of histone H5. *EMBO J.* **15**, 3421–3429 (1996).
- Brown, D. T., Izard, T. & Misteli, T. Mapping the interaction surface of linker histone H1(O) with the nucleosome of native chromatin *in vivo*. *Nature Struct. Mol. Biol.* **13**, 250–255 (2006).
- Chang, X. *et al.* Increased PADI4 expression in blood and tissues of patients with malignant tumors. *BMC Cancer* **9**, 40 (2009).
- Zhang, X. *et al.* Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor α target gene activation. *Proc. Natl Acad. Sci. USA* **109**, 13331–13336 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by programme grants from Cancer Research UK (T.K.) and EMBL (P.B., R.L.). R.P.H.-S. and J.B.G. are supported by the Medical Research Council (G1001690) and the Wellcome Trust. G.C.-B. was funded by EMBO (Long-Term Post-Doctoral Fellowship), European Union (FP7 Marie Curie Intra-European Fellowship for Career Development) and Swedish Research Council. M.A.C. was funded by an EMBO Long-Term Post-Doctoral Fellowship and a Human Frontier Science Programme Long-Term Post-Doctoral Fellowship. C.S.O. was supported by FAPESP (Foundation for Research Support of the State of São Paulo) and mouse embryo work was supported by the Wellcome Trust programme grant to M.Z.-G. M.L.N. was partly supported by the Novo Nordisk Foundation Center for Protein Research, the Lundbeck Foundation, and by and the European Commission's 7th Framework Programme HEALTH-F7-2010-242129/SYBOSS. K.A.M. was funded by NIH grant AI099728. We would like to thank S. Lestari, A. Cook and C. Hill for technical assistance; P. Thompson for providing the TDFA compound; GSK Epinova for CI-amidine; T. Bartke for the gift of histone octamers and help with nucleosome pull-down assays; A. Finch for help with FPLC chromatography; A. Jedrusik for help with embryo work; R. Walker at the Flow Cytometry Core Facility at Wellcome Trust Centre for Stem Cell Research, University of Cambridge and T. Theunissen for help with the flow cytometry; and members of the Kouzarides laboratory for critical discussions of the work. 2TS22C cells were provided by H. Niwa. The ChIP grade H1.2 antibody was a gift from A. Skoultschi.

Author Contributions M.A.C., G.C.-B. and T.K. conceived the idea for this project, designed experiments and wrote the manuscript with the help of all the authors. G.C.-B. and M.A.C. performed ES-cell transductions, established transgenic pre-iPS and ES cell lines and performed gene expression analyses. M.A.C. carried out mutagenesis, protein isolation, biochemical and chromatin immunoprecipitation experiments, and performed citrullination analyses with the help of K.A.M.; G.C.-B. performed reprogramming experiments, with the help of J.S. and A.R.; M.A.C. prepared proteomic samples and M.L.N. performed mass spectrometric analyses. R.P.H.-S. and M.A.C. performed PADI4 treatments of permeabilized cells and subsequent chromatin compaction analyses, with the help of J.B.G. C.S.O. and M.Z.-G. designed and performed mouse embryo experiments. R.L. and P.B. performed bioinformatic analyses of microarray data. T.K. supervised the study.

Author Information Microarray data have been deposited in the ArrayExpress repository (<http://www.ebi.ac.uk/arrayexpress/>) under accession E-MTAB-1975. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.K. (t.kouzarides@gurdon.cam.ac.uk).

Ultraviolet-radiation-induced inflammation promotes angiotropism and metastasis in melanoma

Tobias Bald¹, Thomas Quast², Jennifer Landsberg¹, Meri Rogava¹, Nicole Glodde¹, Dorys Lopez-Ramos¹, Judith Kohlmeyer¹, Stefanie Riesenberger³, Debby van den Boorn-Konijnenberg³, Cornelia Hömig-Hölzel³, Raphael Reuten⁴, Benjamin Schadow⁵, Heike Weighardt⁵, Daniela Wenzel⁶, Iris Helfrich⁷, Dirk Schadendorf⁷, Wilhelm Bloch⁸, Marco E. Bianchi⁹, Claire Lugassy¹⁰, Raymond L. Barnhill¹⁰, Manuel Koch⁴, Bernd K. Fleischmann⁶, Irmgard Förster⁵, Wolfgang Kastenmüller¹¹, Waldemar Kolanus², Michael Hölzel³, Evelyn Gaffal¹ & Thomas Tüting¹

Intermittent intense ultraviolet (UV) exposure represents an important aetiological factor in the development of malignant melanoma¹. The ability of UV radiation to cause tumour-initiating DNA mutations in melanocytes is now firmly established², but how the microenvironmental effects of UV radiation^{3,4} influence melanoma pathogenesis is not fully understood. Here we report that repetitive UV exposure of primary cutaneous melanomas in a genetically engineered mouse model⁵ promotes metastatic progression, independent of its tumour-initiating effects. UV irradiation enhanced the expansion of tumour cells along abluminal blood vessel surfaces and increased the number of lung metastases. This effect depended on the recruitment and activation of neutrophils, initiated by the release of high mobility group box 1 (HMGB1) from UV-damaged epidermal keratinocytes and driven by Toll-like receptor 4 (TLR4). The UV-induced neutrophilic inflammatory response stimulated angiogenesis and promoted the ability of melanoma cells to migrate towards endothelial cells and use selective motility cues on their surfaces. Our results not only reveal how UV irradiation of epidermal keratinocytes is sensed by the innate immune system, but also show that the resulting inflammatory response catalyses reciprocal melanoma-endothelial cell interactions leading to perivascular invasion, a phenomenon originally described as angiotropism in human melanomas by histopathologists⁶. Angiotropism represents a hitherto underappreciated mechanism of metastasis⁷ that also increases the likelihood of intravasation and haematogenous dissemination. Consistent with our findings, ulcerated primary human melanomas with abundant neutrophils and reactive angiogenesis frequently show angiotropism and a high risk for metastases. Our work indicates that targeting the inflammation-induced phenotypic plasticity of melanoma cells and their association with endothelial cells represent rational strategies to specifically interfere with metastatic progression.

Melanocytic nevi exposed to doses of UV light at a level causing sunburn show increased suprabasal location of melanocytes and immune cell infiltrates reminiscent of primary melanomas⁸. Experiments in mice revealed that UV-induced skin inflammatory responses can cause the reactive proliferation and migration of melanocytes⁴. We speculated that these microenvironmental effects of UV irradiation promote melanoma development by stimulating the survival, expansion and dissemination of DNA-damaged melanocytes. To investigate this hypothesis, we used genetically engineered HGF-CDK4(R24C) mice in which deregulated receptor tyrosine kinase signalling due to transgenic overexpression of hepatocyte growth factor (HGF) and impaired cell cycle control

due to an oncogenic CDK4(R24C) mutation in the germline cause the development of invasive and metastatic melanomas as seen in patients⁵.

After two erythral doses of UV light, the skin of adult HGF-CDK4(R24C) mice showed a prominent inflammatory infiltrate consisting predominantly of CD11b⁺Ly6C⁺Ly6G⁺ neutrophils and, to a lesser extent, of CD11b⁺Ly6C⁺Ly6G⁻ inflammatory monocytes and CD11b⁺Ly6C⁻Ly6G⁻ macrophages along with reactive proliferation of epidermal keratinocytes and peripheral blood neutrophilia (Extended Data Fig. 1a–c). Exposure of HGF-CDK4(R24C) mice to UV twice weekly for 6 weeks led to accumulation of melanocytes predominantly in the upper dermis along epidermal basement membranes (Extended Data Fig. 1d). To study the tumour-promoting effects of UV irradiation independent of its tumour-initiating effects, we first initiated melanomas in the back skin of adult HGF-CDK4(R24C) mice with the carcinogen 7,12-dimethylbenz(a)anthracene (DMBA) and then compared tumour development in UV-irradiated and control mice (Fig. 1a). Unexpectedly, we observed that repetitive UV exposure did not affect the incidence, multiplicity and growth kinetics of DMBA-induced primary melanomas in the skin, but rather increased the number of lung metastases (Fig. 1b, c).

In 4 out of 20 UV-irradiated HGF-CDK4(R24C) mice, we observed macroscopically visible expansion of melanoma cells along endothelial cell surfaces of dermal blood vessels (Fig. 1d, e and Extended Data Fig. 2a–c). This angiotropic growth pattern also occurs in about one fifth of patients with human melanomas as an early sign of metastasis (Fig. 1f) and indicates a poor prognosis⁶. Transmission electron microscopy showed melanoma cells covering the abluminal side of endothelial cell surfaces with occasional intravasation into the vessel lumen (Extended Data Fig. 2d). Careful analyses of angiotropism, melanoma-infiltrating immune cells and metastatic spread to draining lymph nodes and lungs (Extended Data Fig. 2e–l) revealed a correlation between UV exposure, neutrophilic inflammation, perivascular invasion and spontaneous lung metastases (Fig. 1g). Interestingly, angiotropic growth was recapitulated in metastases of some UV-irradiated mice (Fig. 1h and Extended Data Fig. 2i, j).

Recently, the TLR3–TRIF (TRIF, also known as TICAM1) pathway was shown to play a role for the recognition of UV-induced alterations in the double-stranded domains of some non-coding RNAs derived from epidermal keratinocytes⁹. We therefore studied the role of innate immune pathways for the induction of skin inflammatory responses following UV irradiation. We observed substantially reduced neutrophilic inflammation and diminished reactive keratinocyte hyperproliferation after two erythral doses of UV in *Tlr4*^{-/-} and *Myd88*^{-/-} mice but not in *Tlr3*^{-/-} and *Trif*^{-/-} mice when compared to wild-type

¹Laboratory of Experimental Dermatology, Department of Dermatology and Allergy, University of Bonn, 53115 Bonn, Germany. ²Molecular Immunology and Cell Biology, Life and Medical Sciences Institute, University of Bonn, 53115 Bonn, Germany. ³Unit for RNA Biology, Department of Clinical Chemistry and Clinical Pharmacology, University of Bonn, 53105 Bonn, Germany. ⁴Institute for Dental Research and Oral Musculoskeletal Biology, Center for Biochemistry, Medical Faculty, University of Cologne, D-50931 Cologne, Germany. ⁵Immunology and Environment, Life and Medical Sciences Institute, University of Bonn, 53115 Bonn, Germany. ⁶Institute for Physiology I, Life & Brain Center, University of Bonn, 53105 Bonn, Germany. ⁷Department of Dermatology, University Hospital Essen, 45122 Essen, Germany. ⁸Institute of Cardiovascular Research and Sport Medicine, Department of Molecular and Cellular Sport Medicine, German Sport University Cologne, 50933 Cologne, Germany. ⁹Division of Genetics and Cell Biology, San Raffaele University and Scientific Institute, 20132 Milan, Italy. ¹⁰Department of Pathology and Laboratory Medicine, Jonsson Comprehensive Cancer Center, University of California Los Angeles (UCLA) Medical Center, Los Angeles, California 90095, USA. ¹¹Institutes of Molecular Medicine and Experimental Immunology, University of Bonn, 53105 Bonn, Germany.

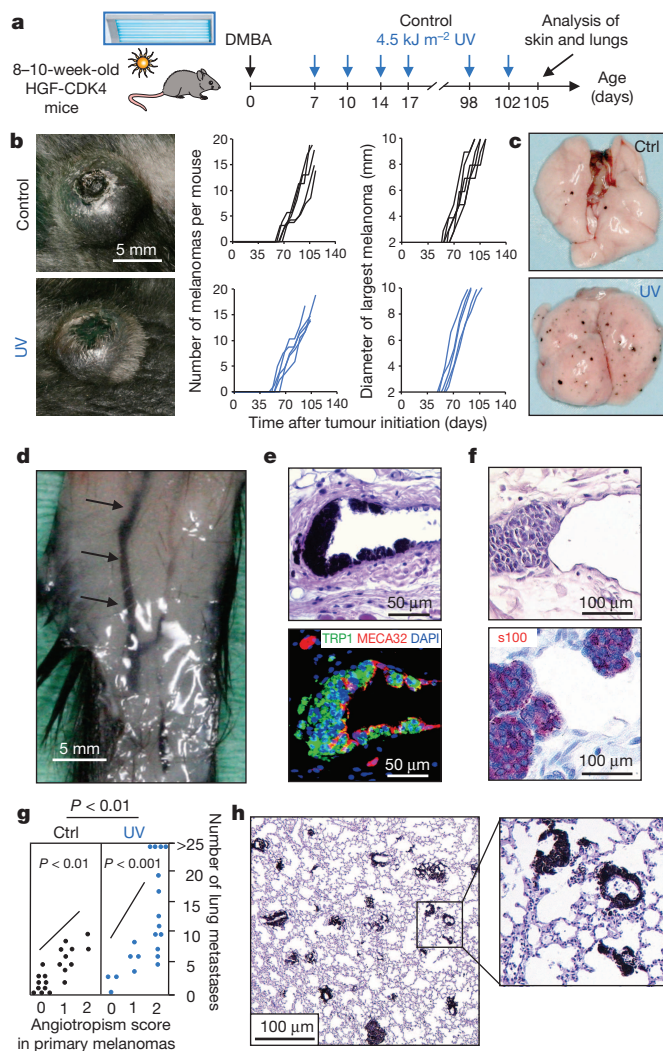


Figure 1 | UV irradiation of skin with DMBA-initiated primary melanomas in HGF-CDK4(R24C) mice selectively promotes angiotropic invasion and lung metastasis. **a**, Experimental model. **b**, Macroscopic appearance (left), mean number (middle) and growth kinetics (right) of DMBA-induced skin melanomas in representative cohorts of 5 UV-exposed and control mice. **c**, Macroscopic appearance of lung metastases. **d**, **e**, Macroscopically visible melanoma cell expansion along dermal blood vessels (arrows) in an UV-irradiated mouse and corresponding haematoxylin and eosin (H&E)-stained section and immunofluorescence analysis. **f**, Angiotropic growth in a human primary melanoma. **g**, Correlation of UV exposure, skin melanoma angiotropism and lung metastasis ($n = 20$; Kruskal-Wallis test with Dunn's correction for multiple comparison). **h**, Angiotropic growth of metastatic melanoma in the lungs of an UV-irradiated mouse.

controls (Fig. 2a, b). Upregulation of *p21* mRNA indicating DNA damage responses was similar in UV-treated mice of all genotypes. The UV-induced induction of skin angiogenesis was also selectively decreased in *Tlr4*^{-/-} but not in *Tlr3*^{-/-} mice compared to controls (Fig. 2c). Experiments with conditional gene knock-in mice expressing MYD88 exclusively in LysM⁺ or K5⁺ cells revealed that UV-induced skin inflammation involves MYD88-dependent TLR4-signalling predominantly in myeloid immune cells and only to a lesser extent in keratinocytes (Fig. 2a, b and Extended Data Fig. 3a–e).

HMGB1 is a ubiquitously expressed nuclear protein that participates in DNA repair¹⁰ and also has TLR4-dependent inflammatory cytokine activities following its release into the extracellular space by severely stressed cells^{11,12}. Upon UV irradiation, we observed cytosolic translocation and release of HMGB1 by epidermal keratinocytes and

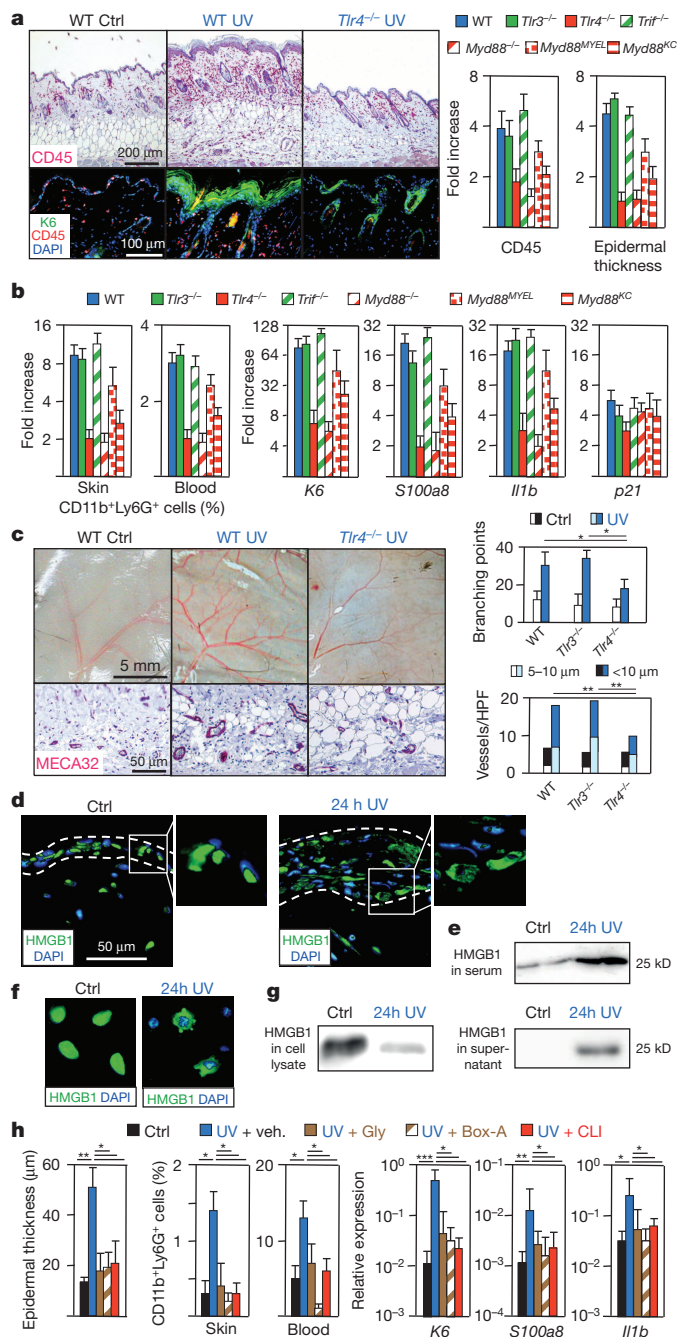


Figure 2 | UV irradiation induces a TLR4/MYD88-driven neutrophilic skin inflammatory response that is initiated by HMGB1 release from UV-damaged keratinocytes. **a**, Differential effect of two UV radiation doses, at a level to cause sunburn, on skin immune cell infiltration and keratinocyte hyperproliferation in wild-type and *Tlr4*^{-/-} mice (left) and relative increases of immune cells and epidermal thickness in mice with the indicated genotypes (right). **b**, Corresponding relative increases of skin and blood neutrophils and skin mRNA levels for the indicated genes. **c**, Macroscopic and microscopic analyses of dermal blood vessels. **d**, **f**, Immunofluorescence analysis for HMGB1 in UV-irradiated and control skin or cultured keratinocytes. **e**, **g**, Corresponding immunoblots for HMGB1 in serum, cell lysates or culture medium. **h**, Effect of treatment with glycyrrhizin (Gly), Box-A, CLI-095 (CLI) or vehicle control on UV-induced keratinocyte proliferation, neutrophil recruitment and skin mRNA levels. All results are representative for 6 mice, error bars indicate the mean \pm s.e.m., * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (unpaired two-tailed Student's *t*-test).

increased HMGB1 serum levels (Fig. 2d, e). HMGB1 cytosolic translocation and release was also found in UV-irradiated primary keratinocyte cultures (Fig. 2f, g). Treatment of mice with recombinant BoxA or glycyrrhizin, two functional inhibitors of extracellular HMGB1, or with CLI-095, a specific inhibitor of TLR4 signalling, largely abrogated UV-dependent neutrophilic inflammation and reactive epidermal keratinocyte hyperproliferation (Fig. 2h). Similar results were obtained after antibody-mediated depletion of Ly6G⁺ neutrophils, whereas clodronate liposome-mediated depletion of macrophages was ineffective (Extended Data Fig. 3f, g). These results provide a HMGB1-dependent mechanistic link between UV-induced DNA damage of epidermal keratinocytes, and a TLR4/MYD88-dependent feed-forward signalling loop predominantly active in neutrophil leukocytes which orchestrates skin inflammation.

To evaluate whether the UV-induced, TLR4/MYD88-driven neutrophilic inflammatory response was also causal for increased angiotropism and metastasis of melanoma, we took advantage of serial HGF-CDK4 (R24C) melanoma skin transplants that spontaneously develop lung

metastases. Repetitive UV exposure of the back skin comprising the tumour transplantation site of HcMel12 melanomas did not affect local growth kinetics but increased angiotropic invasion and lung metastases when compared to non-irradiated controls (Fig. 3a–d and Extended Data Fig. 4a–e). Similar results were obtained with a second, independently generated serial skin transplant HcMel31. The expansion of enhanced green fluorescent protein (EGFP)-expressing HcMel12 cells in the inflamed skin of UV-irradiated mice could be visualized by whole-mount confocal immunofluorescence microscopy two weeks after transplantation (Fig. 3c). This serial transplant model allowed us to study melanoma cell behaviour following UV irradiation in mice with genetic deficiencies of the TLR4/MYD88 signalling axis.

We found that the UV-dependent increase of skin melanoma angiotropism and spontaneous lung metastasis was completely abrogated in *Tlr4*^{−/−} and *Myd88*^{−/−} mice (Fig. 3e and Extended Data Fig. 4f, g). Investigations with conditional *Myd88* knock-in mice revealed that MYD88 signalling was predominantly required in myeloid immune cells (Extended Data Fig. 4h, i). Antibody-mediated depletion experiments confirmed a critical role for Ly6G⁺ neutrophils (Fig. 3f and Extended Data Fig. 4j, k). Importantly, the UV-dependent increase of angiotropism and lung metastases was also abolished by pharmacologic inhibition of extracellular HMGB1 or TLR4 signalling (Fig. 3g and Extended Data Fig. 4l, m). These observations in an independent experimental setting recapitulated key features of our findings with DMBA-induced primary melanomas and unambiguously demonstrated that UV-induced, TLR4/MYD88-driven neutrophilic inflammation promotes angiotropism and metastasis in melanoma.

Next, we studied melanoma-endothelial cell interactions under inflammatory conditions *in vitro*. In transwell assays, we observed that pre-incubation of HcMel12 cells with neutrophil-conditioned medium enhanced their migration towards endothelial cells and, to a lesser extent, towards keratinocytes. This was mimicked by TNF, a key inflammatory cytokine present in UV-induced skin that is also secreted by activated neutrophils (Fig. 4a and Extended Data Fig. 5a, b). Inflammatory activation of HcMel12 cells with TNF also increased their adhesion to vascular extracellular matrix components (Fig. 4b). Time-lapse video microscopy revealed that HcMel12 cells selectively migrated with increased velocities and for longer distances on endothelial cell surfaces when compared to surfaces of keratinocytes or purified extracellular matrix components alone, an effect that was significantly enhanced by TNF (Fig. 4c, Extended Data Fig. 5c and Supplementary Videos 1 and 2). TNF present in neutrophil-conditioned medium also enhanced endothelial cell sprouting from aortic ring explants and promoted the pericyte-like expansion of co-cultured melanoma cells along such endothelial outgrowths (Fig. 4d and Extended Data Fig. 5d–g). Finally, we could demonstrate that melanoma cells preferentially migrated towards, and tightly associated with, blood vessel endothelial cells in inflamed ear tissue explants from UV-irradiated mice (Fig. 4e, Extended Data Fig. 5h–l and Supplementary Videos 3 and 4).

A panel of human melanoma cell lines carrying characteristic *BRAF* or *RAS* mutations also showed enhanced migration towards human vascular endothelial cells in transwell assays and adhered more firmly to vascular extracellular matrix components following inflammatory activation with TNF (Fig. 4f, g and Extended Data Fig. 6a–c). Human melanoma cells also selectively migrated more efficiently on endothelial cell surfaces when compared to surfaces of keratinocytes or extracellular matrix components when activated with TNF (Fig. 4h and Extended Data Fig. 6d, e). Human melanoma cells also invaded inflamed mouse ear tissue explants and closely associated with blood vessel endothelial cells (Fig. 4i and Extended Data Fig. 6f). Collectively, our observations in these *ex vivo* and *in vitro* assays provided evidence that a neutrophilic inflammatory response promoted angiotropic properties of melanoma cells, reflected by the selective migration towards and on endothelial cell surfaces, both in the HGF-CDK4(R24C) mouse model and in human melanoma cells bearing common genetic alterations.

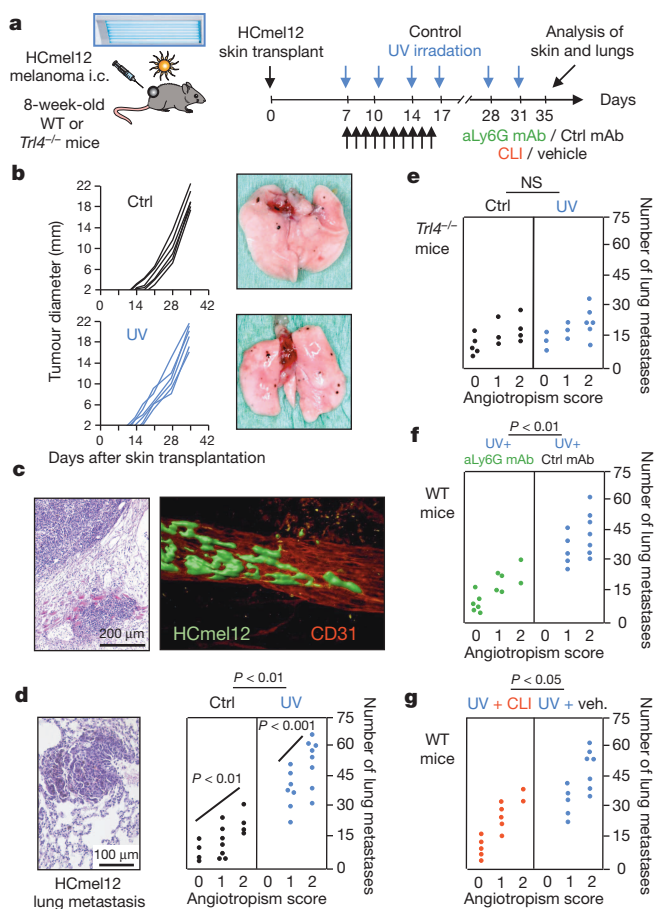


Figure 3 | Increased angiotropism and metastatic dissemination of skin melanomas in UV-irradiated mice depends on TLR4-driven neutrophilic inflammation. **a**, Experimental model. **b**, Growth kinetics of serial HcMel12 melanoma skin transplants in UV-exposed and control mice (*n* = 5, left), macroscopic appearance of lung metastases (right). **c**, Angiotropic growth of EGFP-expressing HcMel12 cells in UV-irradiated skin. H&E stained section (left) and 3D reconstructed whole mount confocal immunofluorescence image (right). **d**, H&E-stained section of a spontaneous HcMel12 lung metastasis (left) and effect of UV exposure on skin melanoma angiotropism and lung metastasis (*n* = 15, right). **e**, Effect of UV irradiation on angiotropism and lung metastasis in TLR4-deficient mice (*n* = 12). **f**, **g**, Effect of antibody-mediated Ly6G⁺ neutrophil depletion and TLR4-inhibition on UV-induced angiotropism and lung metastases (*n* = 12). Cumulative results of 3 independent experiments are shown in d–g (Kruskal–Wallis test with Dunn’s correction for multiple comparison).

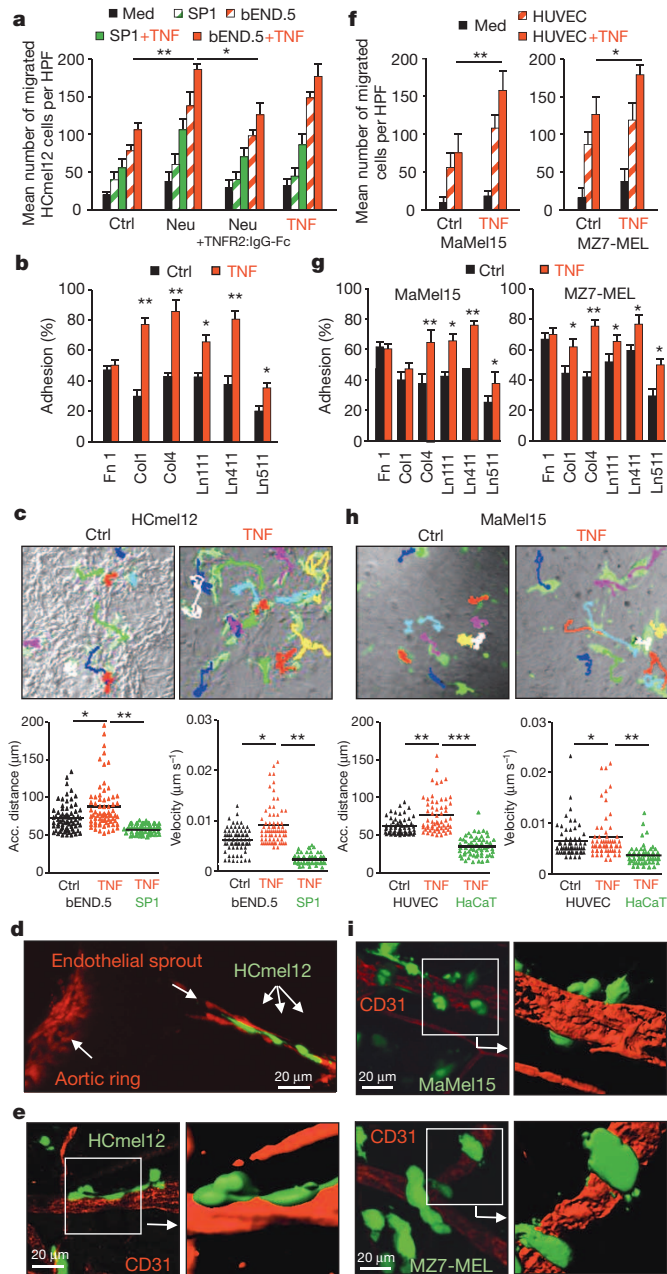


Figure 4 | An inflammatory environment promotes the ability of mouse and human melanoma cells to migrate towards endothelial cells and use selective motility cues on their surfaces. **a**, Migration of HCmel12 melanoma cells towards bEND.5 endothelial cells or SP1 keratinocytes in transwell assays, influence of neutrophil-derived inflammatory mediators (Neu) and role of TNF. Shown are mean number of migrated cells in triplicate determinations \pm s.e.m. **b**, Static adhesion of TNF-activated and control HCmel12 cells to the indicated extracellular matrix components. Shown are mean of triplicate determinations \pm s.e.m. **c**, Two-dimensional (2D) migration tracks of individual control and TNF-activated EGFP-expressing HCmel12 cells on a surface of confluent bEND.5 endothelial cells imaged by time lapse microscopy over 12 h (upper panel), average accumulated distances and migration velocities of HCmel12 cells on bEND.5 endothelial cell or SP1 keratinocyte surfaces (lower panel, $n = 70$ tracks of individual cells, bars indicate the mean). **d**, **e**, Confocal immunofluorescence images showing EGFP-expressing HCmel12 melanoma cells in a pericyte-like association with activated CD31⁺ endothelial cells sprouting from inflamed aortic rings or present in ear tissue explants from UV-irradiated mice. **f**–**i**, Similar experimental results in transwell migration, static adhesion, 2D cell surface migration (on HUVEC vs HaCaT keratinocytes) and ear tissue invasion assays with the indicated human melanoma cell lines. Data in **a** and **f** are representative for 1 out of 5, data in **b**, **c** and **g**, **h** for 1 out of 3 independently performed experiments; statistical significance was determined using the unpaired two-tailed Student's *t*-test for **a**, **b**, **f**, **g** and the unpaired two-tailed Mann–Whitney *U*-test for **c**, **h**; * $P < 0.05$; ** $P < 0.01$.

Overall, we found that neutrophils promote tumour progression, consistent with findings showing that the presence of neutrophils within tumours and their increase in peripheral blood indicate a poor prognosis in melanoma patients^{20,21}. In an unselected sentinel node-staged patient cohort treated at our institution we verified that superficial ulceration and prominent neutrophil infiltration of primary melanomas is associated with signs of angiotropism and lymph node metastasis (Extended Data Fig. 9), confirming published reports of select cases⁶. Our experimental results indicate that neutrophilic inflammation in ulcerated melanomas shifts melanoma cells towards a migratory phenotype, activates endothelial cells and fosters reciprocal melanoma–endothelial interactions, which together promotes angiotropic invasion and metastasis. This pathogenetic connection also explains why adjuvant administration of type I interferons, which inhibit neutrophil function and angiogenesis²², is particularly effective in patients with ulcerated primary melanomas²³.

We show that UV irradiation not only causes tumour-initiating genomic alterations in melanocytes but also promotes their perivascular expansion and metastatic dissemination through TLR4/MYD88-driven neutrophilic inflammation initiated by HMGB1 release from UV-damaged keratinocytes (Extended Data Fig. 10). In addition to TNF, it is probable that other cytokines also enhance angiotropism of melanoma cells, however, dissecting their individual contribution will require further investigations with the respective knockout mice. Our work is in accordance with increasing evidence that HMGB1 and other endogenous TLR ligands critically contribute to tumour-promoting inflammatory networks^{24–27}, and that neutrophils are important players in this process^{22,28,29}. Neutrophils may be particularly important in melanomas which frequently secrete large amounts of CXCR2 ligands²⁸. In addition to their role in primary melanomas shown in this work, neutrophils can also promote tumour–endothelial interactions in the lungs to facilitate metastatic seeding through diverse and complex mechanisms^{27–29}. This includes the recently discovered role of neutrophil extracellular DNA traps, which also activate the coagulation system and probably represent a powerful innate immune stimulus²⁹.

Using immunocompetent HGF-CDK4(R24C) mice in combination with *ex vivo* and *in vitro* assays, here we present pathophysiologically relevant model systems that open up new avenues to further study the complex molecular and cellular mechanisms governing reciprocal melanoma–endothelial cell interactions at the primary tumour site. These experimental tools also provide opportunities to delineate novel

Bioinformatic analyses of public data sets revealed that inflammatory activation of human melanoma cells with TNF induces genes enriched for the GO terms ‘cell migration’, ‘cell adhesion’ and ‘angiogenesis’ (Extended Data Fig. 7 and Supplementary Tables 1 and 2). A subset of melanoma cell lines exhibits a partial baseline activation of this TNF-responsive signature, which includes genes implicated in vasculogenic mimicry¹³ and genes induced upon co-culture with endothelial cells¹⁴. We could confirm these in global transcriptional responses to TNF which strongly correlated in a set of well-characterized BRAF and RAS mutant melanoma cell lines² (Extended Data Fig. 8a, b and Supplementary Table 3). A core set of TNF-regulated genes known to be involved in tumour–endothelial interactions and angiogenesis was further validated by quantitative PCR with reverse transcription (qRT–PCR) in human and mouse melanoma cells (Extended Data Fig. 8c). Together with our experimental observations, these data support the emerging concept that environmental inflammatory signals can increase tumour cell plasticity and promote embryonic cell phenotypes^{15,16}. Indeed, neural crest progenitors of melanocytes migrate along blood vessel surfaces into the skin during embryogenesis^{17–19}.

treatment strategies that impair metastatic tumour progression, including targeted inhibition of selective motility cues or of neutrophilic inflammation driven by HMGB1 and the TLR4/MYD88 signalling axis³⁰.

METHODS SUMMARY

Animal experiments and UV irradiation. Melanoma-prone HGF-CDK4(R24C) mice in the C57BL/6 background were bred as described previously¹⁶. HGF-CDK4(R24C) mice with carcinogen-initiated primary melanomas or C57BL/6 wild-type as well as *Tlr4*^{-/-}, *Myd88*^{-/-} and *Myd88*^{MYEL} mice bearing serial melanoma skin transplants were exposed to 4.5 kJ m⁻² UV twice weekly and tumour development was monitored. In some experiments myeloid immune cells were depleted or the HMGB1/TLR4 signalling axis was blocked pharmacologically. All animal experiments were approved by the local government authorities (LANUV, NRW, Germany) and were performed according to the institutional and national guidelines for the care and use of laboratory animals.

Histology, immunostaining, electron microscopy, flow cytometry and RT-PCR. Angiotropism of melanomas, skin inflammatory responses and metastatic dissemination were analysed by H&E staining. Immunostaining was performed with antibodies for melanoma cells, activated keratinocytes, endothelial cells and immune cells. Ultrastructural analyses were performed with transmission electron microscopy. The composition of immune cells in tissues and blood was determined by flow cytometry. Gene expression was quantified by real-time RT-PCR analyses. Primers are given in Supplementary Table 4.

Organ and cell culture experiments. Primary keratinocytes were cultured from newborn skin, irradiated with UV and the release of HMGB1 determined by immunofluorescence and immunoblot analyses. The effect of pro-inflammatory mediators on the migratory behaviour of melanoma cells towards and on endothelial cell surfaces was investigated in transwell and novel two-dimensional cell surface migration assays *in vitro*. Melanoma-endothelial cell interactions were further studied in co-cultures of EGFP-expressing tumour cells with red fluorescent aortic ring explants. An ear tissue invasion assay was established in which fluorescent melanoma cells were seeded onto ear tissue explants from UV-irradiated or control mice and allowed to migrate into the dermis. Ear tissue whole mounts were immunostained for endothelial cells and imaged with confocal laser-scanning immunofluorescence microscopy.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 May 2013; accepted 29 January 2014.

Published online 26 February 2014.

- Chang, Y. M. *et al.* Sun exposure and melanoma risk at different latitudes: a pooled analysis of 5700 cases and 7216 controls. *Int. J. Epidemiol.* **38**, 814–830 (2009).
- Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251–263 (2012).
- Haass, N. K. & Herlyn, M. Normal human melanocyte homeostasis as a paradigm for understanding melanoma. *J. Invest. Dermatol. Symp. Proc.* **10**, 153–163 (2005).
- Zaidi, M. R. *et al.* Interferon- γ links ultraviolet radiation to melanomagenesis in mice. *Nature* **469**, 548–553 (2011).
- Gaffal, E. *et al.* Neonatal UVB exposure accelerates melanoma growth and enhances distant metastases in Hgf-Cdk4^{R24C} C57BL/6 mice. *Int. J. Cancer* **129**, 285–294 (2011).
- Barnhill, R. L. & Lugassy, C. Angiotropic malignant melanoma and extravascular migratory metastasis: description of 36 cases with emphasis on a new mechanism of tumour spread. *Pathology* **36**, 485–490 (2004).
- Valastyan, S. & Weinberg, R. A. Tumour metastasis: molecular insights and evolving paradigms. *Cell* **147**, 275–292 (2011).
- Tronnier, M., Smolle, J. & Wolff, H. H. Ultraviolet irradiation induces acute changes in melanocytic nevi. *J. Invest. Dermatol.* **104**, 475–478 (1995).
- Bernard, J. J. *et al.* Ultraviolet radiation damages self noncoding RNA and is detected by TLR3. *Nature Med.* **18**, 1286–1290 (2012).
- Lange, S. S., Mitchell, D. L. & Vasquez, K. M. High mobility group protein B1 enhances DNA repair and chromatin modification after DNA damage. *Proc. Natl Acad. Sci. USA* **105**, 10320–10325 (2008).
- Yang, H. *et al.* A critical cysteine is required for HMGB1 binding to Toll-like receptor 4 and activation of macrophage cytokine release. *Proc. Natl Acad. Sci. USA* **107**, 11942–11947 (2010).
- Venereau, E. *et al.* Mutually exclusive redox forms of HMGB1 promote cell recruitment or proinflammatory cytokine release. *J. Exp. Med.* **209**, 1519–1528 (2012).
- Kirschmann, D. A., Seftor, E. A., Hardy, K. M., Seftor, R. E. & Hendrix, M. J. Molecular pathways: vasculogenic mimicry in tumor cells: diagnostic and therapeutic implications. *Clin. Cancer Res.* **18**, 2726–2732 (2012).

- Lugassy, C. *et al.* Pilot study on “pericytic mimicry” and potential embryonic/stem cell properties of angiotropic melanoma cells interacting with the abluminal vascular surface. *Cancer Microenviron.* **6**, 19–29 (2013).
- Schwitalla, S. *et al.* Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38 (2013).
- Landsberg, J. *et al.* Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. *Nature* **490**, 412–416 (2012).
- Nagy, N. *et al.* Endothelial cells promote migration and proliferation of enteric neural crest cells via β 1 integrin signaling. *Dev. Biol.* **330**, 263–272 (2009).
- Gupta, P. B. *et al.* The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. *Nature Genet.* **37**, 1047–1054 (2005).
- Lugassy, C., Peault, B., Wadehra, M., Kleinman, H. K. & Barnhill, R. L. Could pericytic mimicry represent another type of melanoma cell plasticity with embryonic properties? *Pigment Cell Melanoma Res.* **26**, 746–754 (2013).
- Jensen, T. O. *et al.* Intratumoral neutrophils and plasmacytoid dendritic cells indicate poor prognosis and are associated with pSTAT3 expression in AJCC stage I/II melanoma. *Cancer* **118**, 2476–2485 (2012).
- Schmidt, H. *et al.* Pretreatment levels of peripheral neutrophils and leukocytes as independent predictors of overall survival in patients with American Joint Committee on Cancer Stage IV Melanoma: results of the EORTC 18951 Biochemotherapy Trial. *J. Clin. Oncol.* **25**, 1562–1569 (2007).
- Jablonska, J., Leschner, S., Westphal, K., Lienenklaus, S. & Weiss, S. Neutrophils responsive to endogenous IFN- β regulate tumor angiogenesis and growth in a mouse tumor model. *J. Clin. Invest.* **120**, 1151–1164 (2010).
- Eggermont, A. M. *et al.* Ulceration and stage are predictive of interferon efficacy in melanoma: results of the phase III adjuvant trials EORTC 18952 and EORTC 18991. *Eur. J. Cancer* **48**, 218–225 (2012).
- Mittal, D. *et al.* TLR4-mediated skin carcinogenesis is dependent on immune and radioresistant cells. *EMBO J.* **29**, 2242–2252 (2010).
- West, X. Z. *et al.* Oxidative stress induces angiogenesis by activating TLR2 with novel endogenous ligands. *Nature* **467**, 972–976 (2010).
- Kim, S. *et al.* Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. *Nature* **457**, 102–106 (2009).
- Hiratsuka, S. *et al.* Primary tumours modulate innate immune signalling to create pre-metastatic vascular hyperpermeability foci. *Nature Commun.* **4**, 1853 (2013).
- Huh, S. J., Liang, S., Sharma, A., Dong, C. & Robertson, G. P. Transiently entrapped circulating tumor cells interact with neutrophils to facilitate lung metastasis development. *Cancer Res.* **70**, 6071–6082 (2010).
- Cools-Lartigue, J. *et al.* Neutrophil extracellular traps sequester circulating tumor cells and promote metastasis. *J. Clin. Invest.* **123**, 3446–3458 (2013).
- Guo, Z. S., Liu, Z., Bartlett, D. L., Tang, D. & Lotze, M. T. Life after death: targeting high mobility group box 1 in emergent cancer therapies. *Am. J. Cancer Res.* **3**, 1–20 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank the following individuals for their support: G. Merlino, M. Barbacid, A. Nagy, M. Gertsenstein and J. Takeda for generously providing genetically engineered mice; S. Bald, A. Spörle, C. Lemke, T. Artz and P. Aymans for managing the mouse colony and performing tumour analyses. This research was funded in part by the following grants: Deutsche Krebshilfe P9 in the Melanoma Research Network and DFG A12 in the SFB832 as well as A22 in the SFB704 to T.T., BONFOR to J.L., J.K. and C.H.-H., DFG SFB832 core support and HO 4281/2-1 to M.H., DFG A7 in the SFB832 and A2 in the SFB829 to M.K., DFG A8 in the SFB704 to W. Ko., DFG A1 in the SFB704 to I.F., DFG SFB829 Z2 to W.B., NRW junior research group to D.W., Jürgen Manchot Stiftung to H.W. and AIRC to M.E.B. Additionally, the DFG supports I.F., W.Ka., W.Ko., M.H. and T.T. as members of the Excellence Cluster ImmunoSensation.

Author Contributions T.B. coordinated and performed all animal experiments and tissue analyses. T.B. and T.Q. developed and performed transwell, 2D cell surface migration and ear tissue invasion assays. M.R. performed keratinocyte culture and melanoma-endothelial cell co-culture in aortic ring assays. J.L., N.G. and D.L.-R. helped with animal experiments and tissue analyses. J.L. and J.K. collected and analysed patient data. R.R. and M.K. performed static adhesion assays. D.W. and B.K.F. helped to design and interpret aortic ring assay experiments. W.Ka. helped with whole mount immunostaining, confocal microscopy and image analysis. W.B. performed transmission electron microscopy. M.R., J.K., S.R., D.v.d.B.-K. and C.H.-H. performed experiments with human cell lines. I.H. and D.S. provided human melanoma cell lines, helped to design experiments and interpret the data. M.E.B. helped to design and interpret experiments involving HMGB1. C.L. and R.B. helped to interpret angiotropism in mice and human histopathology. B.S., H.W. and I.F. helped to design and interpret experiments involving *Myd88*^{LSL} mice. W.Ko. helped to design and interpret migration assays. M.H. coordinated human melanoma cell culture and performed microarray experiments and bioinformatic analyses. E.G. helped to design and coordinate all animal experiments. T.B., T.Q., M.R. and J.L. helped to prepare figures and methods. W.Ko., M.H. and E.G. interpreted data and helped to write the manuscript. T.T. conceived and supervised all aspects of the project, designed experiments, interpreted the data and wrote the manuscript. All authors reviewed the manuscript.

Author Information The microarray data have been deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE51221. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H. (michael.hoelzel@ukb.uni-bonn.de), E.G. (evelyn.gaffal@ukb.uni-bonn.de) or T.T. (thomas.tueting@ukb.uni-bonn.de).

Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle

Karin Denger¹, Michael Weiss², Ann-Katrin Felux², Alexander Schneider³, Christoph Mayer³, Dieter Spiteller¹, Thomas Huhn⁴, Alasdair M. Cook¹ & David Schleheck¹

Sulphoquinovose (SQ, 6-deoxy-6-sulphoglucose) has been known for 50 years as the polar headgroup of the plant sulfolipid^{1,2} in the photosynthetic membranes of all higher plants, mosses, ferns, algae and most photosynthetic bacteria³. It is also found in some non-photosynthetic bacteria⁴, and SQ is part of the surface layer of some Archaea⁵. The estimated annual production of SQ⁴ is 10,000,000,000 tonnes (10 petagrams), thus it comprises a major portion of the organo-sulphur in nature, where SQ is degraded by bacteria^{6,7}. However, despite evidence for at least three different degradative pathways in bacteria^{6–8}, no enzymic reaction or gene in any pathway has been defined, although a sulphoglycolytic pathway has been proposed⁷. Here we show that *Escherichia coli* K-12, the most widely studied prokaryotic model organism, performs sulphoglycolysis, in addition to standard glycolysis. SQ is catabolised through four newly discovered reactions that we established using purified, heterologously expressed enzymes: SQ isomerase, 6-deoxy-6-sulphofructose (SF) kinase, 6-deoxy-6-sulphofructose-1-phosphate (SFP) aldolase, and 3-sulpholactaldehyde (SLA) reductase. The enzymes are encoded in a ten-gene cluster, which probably also encodes regulation, transport and degradation of the whole sulfolipid; the gene cluster is present in almost all (>91%) available *E. coli* genomes, and is widespread in Enterobacteriaceae. The pathway yields dihydroxyacetone phosphate (DHAP), which powers energy conservation and growth of *E. coli*, and the sulphonate product 2,3-dihydroxypropane-1-sulphonate (DHPS), which is excreted. DHPS is mineralized by other bacteria, thus closing the sulphur cycle within a bacterial community.

Recent work showed that environmental isolates of *Klebsiella* spp. (Enterobacteriaceae) convert SQ quantitatively to DHPS^{7,8}, and we proposed that utilization of SQ might be a property of Enterobacteriaceae. We found that four genome-sequenced *E. coli* K-12 substrains (BW25113, DH1, MG1655 and W3100), after subculturing, grew with SQ within 1 to 3 days. We chose to work (largely) with the fastest-growing substrain, MG1655. The organism used SQ as a sole source of carbon and energy with a molar-growth yield of 3 g of protein per mol of SQ carbon, whereas glucose gave approximately 6 g of protein per mol of carbon; the latter value represented mass balance of glucose carbon as biomass and CO₂ (ref. 9). However, approximately 1 mol of DHPS per mol of SQ was released into the growth medium (Fig. 1a), as observed with *Klebsiella oxytoca*⁸. Thus, there was complete mass balance for carbon and for sulphur from SQ. The growth rate with SQ was 0.13 h⁻¹ (0.5 h⁻¹ with glucose), and the specific degradation rate for SQ *in vivo* was 120 mU per mg of protein (1 mU = 1 nmol min⁻¹). We concluded that SQ is metabolized to a C₃ sulphonate, which is excreted as DHPS, and that the remainder of the molecule is used for growth (Fig. 2a).

The out-grown culture was filter-sterilized and inoculated with *Cupriavidus pinatubonensis* JMP134, which can utilize DHPS for growth¹⁰, but cannot utilize SQ⁸. *C. pinatubonensis* grew with the DHPS formed from SQ by *E. coli*, and released its sulphonate sulphur quantitatively as sulphate (Fig. 1b) using a pathway described elsewhere¹⁰. We thus demonstrated mineralization of SQ in a laboratory model system.

Proteins from whole cells of *E. coli* K-12 grown with glucose or SQ were subjected to two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (Extended Data Fig. 1) and examined by peptide fingerprinting–mass spectrometry (PF-MS) (Extended Data Table 1). The immediately relevant, apparently SQ-inducible proteins (see Extended Data Fig. 1 and Extended Data Table 1) were attributed to b3878 (also known as *yihQ* (b numbers are locus tags); predicted to be an α -glucosidase), b3879 (also known as *yihR*; predicted to be an epimerase), b3880 (*yihS*; predicted to be an isomerase), b3881 (*yihT*; predicted to be an aldolase) and b3882 (*yihU*; predicted to be an NAD⁺/NADH-linked dehydrogenase/reductase). Transcriptional analyses for the gene cluster b3879–b3882, as well as for b3883 (also known as *yihV*; predicted to be a sugar kinase), confirmed a strong inducible transcription during

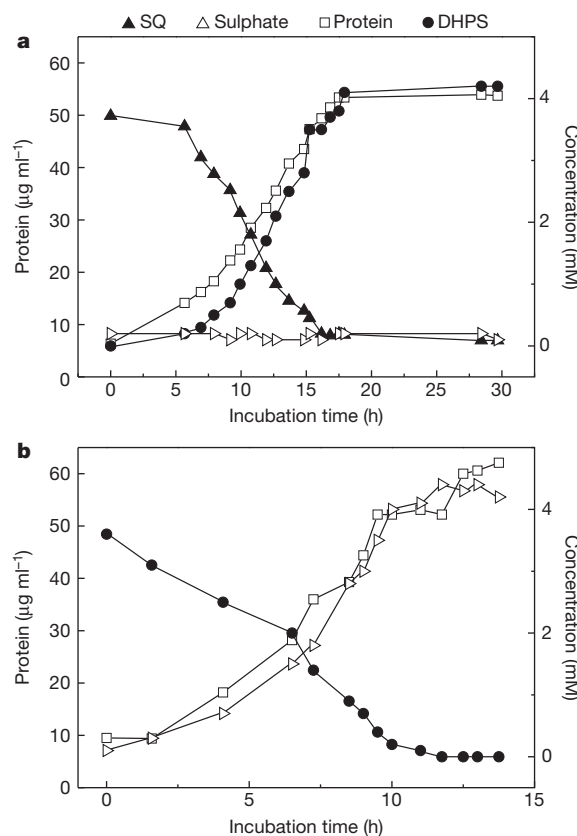


Figure 1 | Complete degradation of sulphoquinovose during growth. a, Growth of *E. coli* K-12 substrain MG1655 with SQ and excretion of 2,3-dihydroxypropane-1-sulphonate (DHPS). b, Growth of *C. pinatubonensis* JMP134 with the DHPS formed from SQ by *E. coli*. Data from representative growth experiments ($n = 3$) are shown. To allow a compact graph, sulphate release and not total sulphate is shown.

¹Department of Biology, University of Konstanz, D-78457 Konstanz, Germany. ²Konstanz Research School Chemical Biology, University of Konstanz, D-78457 Konstanz, Germany. ³Interfaculty Institute of Microbiology and Infection Medicine, University of Tübingen, D-72076 Tübingen, Germany. ⁴Department of Chemistry, University of Konstanz, D-78457 Konstanz, Germany.

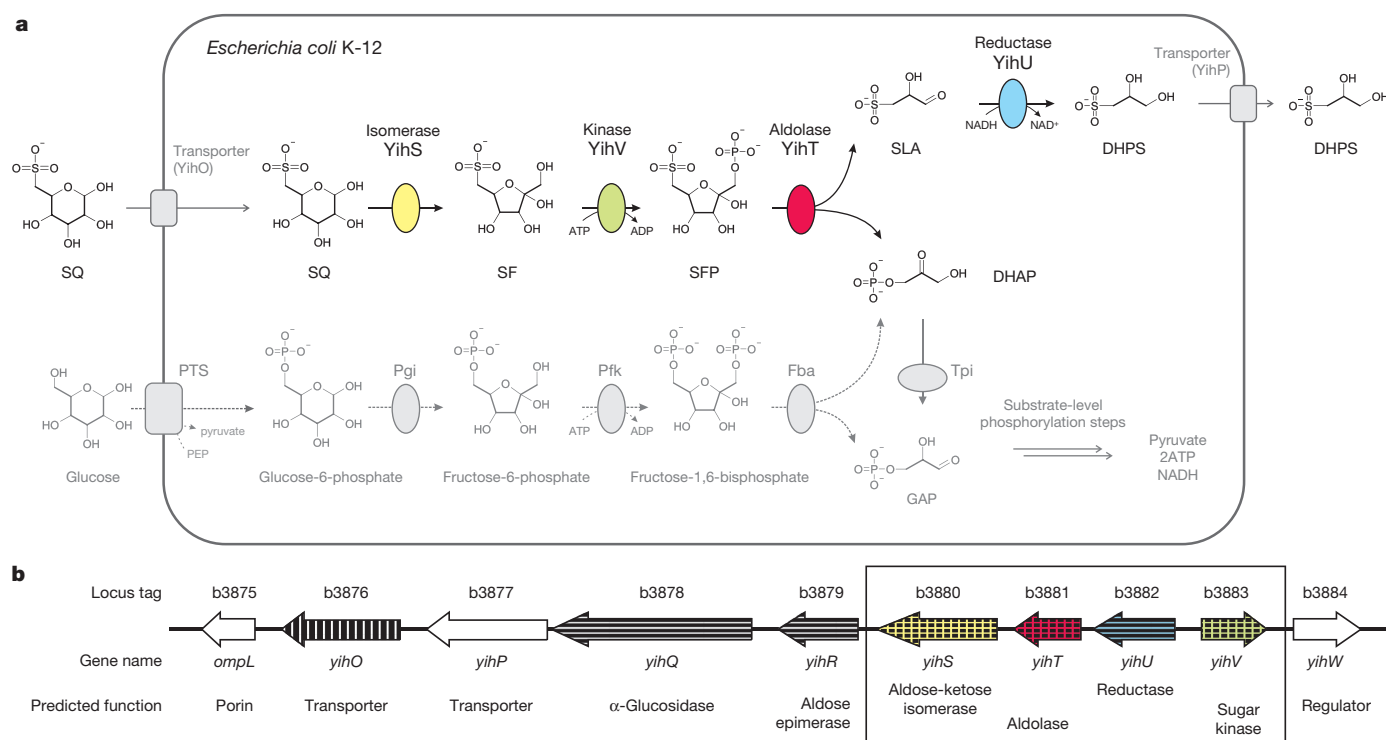


Figure 2 | The four core enzyme reactions of sulphoglycolysis, with transport, and the corresponding genes in a ten-gene cluster in *E. coli* K-12. **a, SQ is metabolized by four enzymes (shown in colour) to a C_3 sulphonate, DHPS, which is excreted, and the remainder of the molecule is used for growth. For comparison, the analogous enzyme reactions for the catabolism of (unsubstituted) glucose through the glycolytic pathway in *E. coli* are also shown (dashed arrows). Fba, fructose biphosphate aldolase; GAP, glyceraldehyde-3-phosphate; Pfk, phosphofructokinase; Pgi, phosphoglucose isomerase; PTS, phosphotransferase system permease;**

Tpi, triose phosphate isomerase. **b**, The EcoGene *E. coli* website (<http://www.EcoGene.org>) uses the abbreviation *yih* for most of these genes; we have retained this nomenclature. Vertical stripes, genes confirmed as being essential for growth with SQ by mutational analysis; horizontal stripes, genes confirmed as being inducible for growth with SQ by proteomic and/or transcriptional analyses; box-framed genes, genes encoding the four core enzymes of the pathway (shown in **a**) that were subject of heterologous expression, purification and functional characterization.

growth with SQ, but not during growth with glucose (Extended Data Fig. 2). Furthermore, single-gene knockouts (in substrain BW25113 (ref. 11)) in genes b3876 (also known as *yihO*; predicted to be a major facilitator superfamily (MFS)-type transporter), b3880, b3881 and b3883 did not grow with SQ, which confirmed and expanded on the proteomic and transcriptional data (Fig. 2b).

We thus identified a gene cluster in *E. coli* K-12 that contained SQ-inducible, essential genes for catabolism of SQ, but we still did not know which pathway was involved. A sulphoglycolytic pathway would involve a hypothetical 3-sulpholactaldehyde (SLA) reductase to yield DHPS in the final reaction (apart from export) (Figs 1a and 2a), whereas a hypothetical Entner–Doudoroff-type or pentose–phosphate-type pathways, or another novel pathway. An SLA reductase was detected (assayed as DHPS oxidation) in cell-free extracts of SQ-grown substrain MG1655 at a specific activity of 420 mU per mg of protein, which exceeds the specific degradation rate for SQ *in vivo* and, thus, was sufficient to explain growth. This enzyme activity was not detected in extracts of glucose-grown cells. Thus the enzyme was confirmed to be inducible, and it was specific for NAD^+ ; $NADP^+$ was not a substrate. Furthermore, SQ did not lead to reduction of NAD^+ or of $NADP^+$ in the extracts of SQ- or glucose-grown cells, hence, hypothetical SQ dehydrogenase was not detectable. These data led us to predict the sulphoglycolytic pathway depicted in Fig. 2a, including the requirement for sulphonate import and export across the cell membrane^{12–14}.

The four predicted core enzymes of the pathway (Fig. 2a) were heterologously expressed and purified as His-tagged proteins, b3880 (putative isomerase), b3883 (putative sugar kinase), b3881 (putative aldolase) and b3882 (putative reductase) (Extended Data Fig. 3). Protein b3882 was shown to encode SLA reductase. First, we partially purified and

identified (PF–MS) the wild-type enzyme in cell-free extracts of substrain MG1655 (see above), and second, we examined the recombinant protein (see below). In both cases we identified that b3882 represents an SLA reductase; the enzyme showed no activity with 4-hydroxybutyrate¹⁵.

The heterologously expressed and purified putative isomerase (b3880) caused about one-fourth of the SQ in the reaction mixture to disappear, as observed by high-pressure liquid chromatography–mass spectrometry (HPLC–MS), and a new peak was formed that eluted with shorter retention time, but exhibited the same relative mass ($M_r = 244$ Dalton (Da); observed as a quasi-molecular ion in the negative ion mode ($[M-H]^-$) at a mass-to-charge ratio (m/z) of 243) (Fig. 3a, b). The new peak was confirmed to represent 6-deoxy-6-sulphofructose (SF), as proposed elsewhere⁷, by the HPLC separation pattern (Extended Data Fig. 4), by the matching exact mass of the $[M-H]^-$ ion (Extended Data Fig. 4), and by its MS–MS fragmentation pattern (Extended Data Fig. 5). Thus we confirmed that b3880 catalysed the SQ isomerase reaction.

The reaction mixture was augmented with ATP and the putative sugar kinase (b3883). The peaks of SQ and SF partially disappeared and a new peak was formed (Fig. 3c). This new peak was confirmed to represent 6-deoxy-6-sulphofructose-1-phosphate (SFP), as proposed elsewhere⁷, by the matching exact mass of the $[M-H]^-$ ion (observed mass, 322.9877 Da; theoretical mass of $C_6H_{12}O_{11}PS^-$, 322.9843 Da) and by its fragmentation pattern (Extended Data Fig. 6). HPLC confirmed that ATP disappeared and ADP was formed during the reaction and, furthermore, that SFP was converted back to SF when alkaline phosphatase was added to a preparation of SFP (not shown). Thus, with b3883, we expressed an ATP-dependent kinase that phosphorylated SF to SFP.

The reaction mixture was augmented with the putative aldolase (b3881). The peak for SFP partially disappeared, and two new peaks were formed

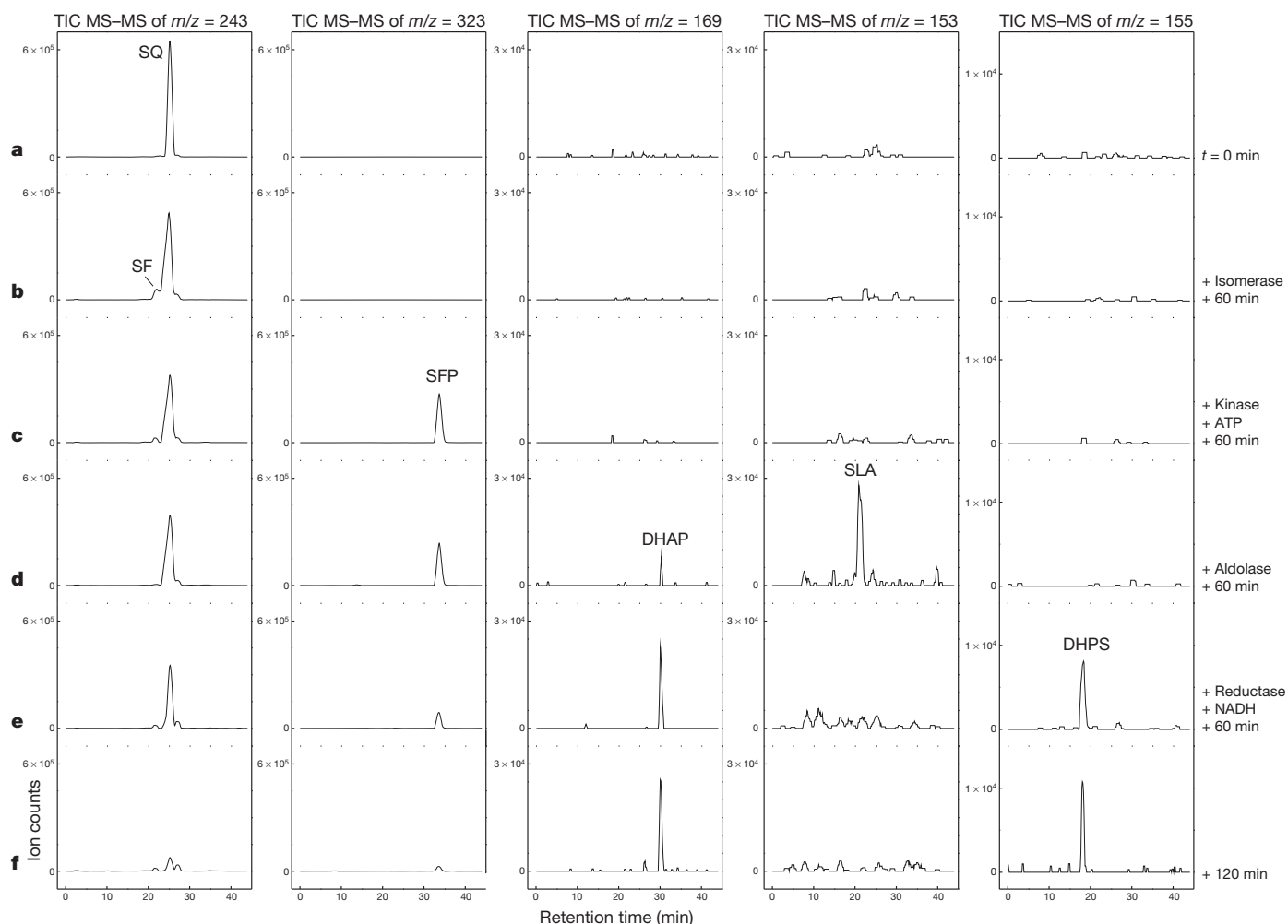


Figure 3 | Illustration of the reactions of the four core enzymes of sulphoglycolysis *in vitro*. The transformation of SQ to SF, SFP, DHAP and SLA, and DHPS, by successive addition of recombinantly expressed pathway enzymes was followed by HPLC–ESI–MS. **a**, Sample of SQ in reaction buffer ($t = 0$ min). **b**, Sample after addition of isomerase (b3880) ($t = 60$ min). **c**, Sample after addition of ATP and kinase (b3883) ($t = 120$ min). **d**, Sample after addition of aldolase (b3881) ($t = 180$ min). **e**, Sample after addition of

(Fig. 3d). The first new peak was identified to represent DHAP, as proposed elsewhere⁷, with an authentic DHAP standard. The second new peak was confirmed to represent SLA, as proposed elsewhere⁷, by the matching mass of the $[M-H]^-$ ion ($M_r = 154$ Da; observed as $[M-H]^-$ ion at $m/z = 153$) and by its fragmentation pattern (Extended Data Fig. 7); the same peak was observed when we used recombinant SLA reductase in reverse to oxidize DHPS to SLA (see above). Thus, with b3881, we expressed an aldolase that cleaved SFP into DHAP and SLA. The SFP turnover was incomplete (Fig. 3d); the equilibrium of the corresponding enzyme reaction in glycolysis (fructose-1,6-bisphosphate aldolase) lies far to the left¹⁶, that is, hardly any products are formed.

However, when NADH and the recombinant SLA reductase (b3882) were added, the peak for SFP was further diminished, as was the peak for SQ, and that for the aldolase-reaction product DHAP was further increased (Fig. 3e). In addition, the peak for SLA had disappeared, and the peak for the anticipated sulphonate product, DHPS, was formed (Fig. 3e and Extended Data Fig. 7). After an extended incubation of the four-enzyme reaction (see Fig. 3e, f), the peaks for SQ and SFP had almost completely disappeared, and the peaks for DHAP and DHPS had further increased.

Together, the results show that the SQ pathway in *E. coli* K-12 (Fig. 2a) does not involve a desulphonation reaction and that no substrate-level phosphorylation of the sulphonated C_3 intermediate occurs, which has

NADH and reductase (b3882) ($t = 240$ min). **f**, Sample after extended incubation of the four-enzyme reaction ($t = 360$ min). The total-ion chromatograms (TICs) recorded in the negative-ion mode from the MS–MS fragmentation of the quasi-molecular ions $[M-H]^-$ of SQ and SF, SFP, DHAP, SLA and DHPS, from a representative experiment ($n = 5$) are shown. For representative MS–MS fragmentation patterns of the $[M-H]^-$ ions of SQ and SF, SFP, and SLA and DHPS, see Extended Data Figs 5, 6 and 7, respectively.

been used previously⁷ as a default hypothesis. Furthermore, we deduce that there are ten genes in the gene cluster (Fig. 2b). The core pathway comprises a SQ transporter (for example, b3876, YihO), SQ isomerase (b3880, YihS), SF kinase (b3883, YihV), SFP aldolase (b3881, YihT), SLA reductase (b3882, YihU) and a DHPS exporter (for example, b3877, YihP), which could be under the putative control of repressor b3884 (YihW). We propose a sulpholipid porin (b3875, OmpL), a sulpholipid α -glucosidase (b3878, YihQ), and an epimerase (b3879, YihR) to funnel other SQ derivatives into the pathway; for example, the whole sulpholipid (see Extended Data Fig. 8).

The gene cluster is found in at least 1,009 (>91%) of the 1,110 genome sequences of commensal *E. coli*, as well as pathogenic *E. coli* (for example, EHEC) strains, that were available in November 2013 (finished and draft genome sequences) in the Integrated Microbial Genomes (IMG) and Human Microbiome Project (HMP) databases (that is, gene clusters with syntenic *yihTUVW* and collinear homologues of *yihSRQPO* and *ompL* in variable order). Hence, the gene cluster is a feature of the core-genome of *E. coli* species. It can also be found in a wide range of other Enterobacteriaceae (for example, *Chronobacter sakazakii* ATCC BAA-894, *Klebsiella oxytoca* 10-5242, *Pantoea ananatis* LMG 20103 and *Salmonella enterica* LT2). We therefore suspect that the pathway has a significant role in bacteria in the alimentary tract of all omnivores and herbivores, that the pathway occurs in bacteria in excrement from

these animals, and that it occurs in plant pathogens, which would explain part of the widespread occurrence of microbial degradation observed^{6–8}.

SQ is produced in large amounts in nature and thus represents a significant proportion of the organic sulphur cycle⁴, and it is degraded in similar amounts by both bacteria^{6–8} and algae¹⁷, or it would accumulate in the environment. We see here that the Enterobacteriaceae use one pathway (Fig. 2a) to initiate degradation of SQ, and that a community is required for complete degradation (Fig. 1b)⁸. This covers a variety of habitats, but we know that other pathways exist. A previous paper⁷ presented evidence for SQ dehydrogenase, which we also observe in our SQ-using strain of *Pseudomonas putida* (A.-K.F. unpublished observations). Notably, another group⁶ reported complete SQ degradation, including desulphonation, in a single organism; however, this organism has been lost¹⁸.

In summary, we have established that sulphoglycolysis, which was named but not defined in a previous report¹⁹, converts SQ to DHPS in the most widely studied prokaryotic model organism, *E. coli* K-12, representing many Enterobacteriaceae (Fig. 2a). We have identified a gene cluster in *E. coli* K-12 (Fig. 2b) that encodes the pathway. The core pathway, for SQ, involves four newly discovered enzymes, two newly identified transporters and three newly characterized intermediates (Fig. 2a, b). We know that the pathway is regulated (Extended Data Figs 1 and 2) and we suspect that it includes the degradation of the whole sulpholipid (Extended Data Fig. 8). The pathway represents a substantial part of the biogeochemical sulphur cycle, and the pathway is likely to have a significant role in bacteria in the alimentary tract of all omnivores and herbivores, and in plant pathogens. We and others^{6–8} anticipate other degradative pathways for SQ in nature; for example, in bacteria of all marine, freshwater and terrestrial habitats where SQ is produced and degraded. We now provide the tools to elucidate these degradative pathways.

METHODS SUMMARY

SQ and DHPS were synthesized chemically and identified by NMR and mass spectrometry^{8,10}. Cultivation, preparation of cell-free extracts, enzyme purification, 2D-PAGE and PF-MS, RNA preparation and polymerase chain reaction with reverse transcription (RT-PCR), and expression and purification of His-tagged proteins, are described in the online Methods. SQ, SF, SFP, SLA, DHAP and DHPS were separated using hydrophilic interaction liquid chromatography (HILIC)⁸ and detected by an evaporative light scattering detector (ELSD)⁸ or electrospray ionization (ESI)-time-of-flight (TOF)-MS or ESI-iontrap-MS (see Methods). The enzyme reaction mixture (Fig. 3) was 3 mM SQ in 50 mM ammonium acetate buffer (pH 8.7), and 8 mM ATP, 0.5 mM MgCl₂ and 8 mM NADH supplemented with the corresponding enzymes (each 50 µg ml⁻¹).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 August; accepted 12 December 2013.

Published online 26 January 2014.

1. Benson, A. A. The plant sulfolipid. *Adv. Lipid Res.* **1**, 387–394 (1963).

2. Benning, C. Questions remaining in sulfolipid biosynthesis: a historical perspective. *Photosynth. Res.* **92**, 199–203 (2007).
3. Benning, C. Biosynthesis and function of the sulfolipid sulfoquinovosyl diacylglycerol. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 53–75 (1998).
4. Harwood, J. L. & Nicholls, R. G. The plant sulfolipid — a major component of the sulphur cycle. *Biochem. Soc. Trans.* **7**, 440–447 (1979).
5. Meyer, B. H. et al. Sulfoquinovose synthase — an important enzyme in the N-glycosylation pathway of *Sulfolobus acidocaldarius*. *Mol. Microbiol.* **82**, 1150–1163 (2011).
6. Martelli, H. L. Oxidation of sulphonic compounds by aquatic bacteria isolated from rivers of the Amazon region. *Nature* **216**, 1238–1239 (1967).
7. Roy, A. B., Hewlins, M. J. E., Ellis, A. J., Harwood, J. L. & White, G. F. Glycolytic breakdown of sulfoquinovose in bacteria: a missing link in the sulfur cycle. *Appl. Environ. Microbiol.* **69**, 6434–6441 (2003).
8. Denger, K., Huhn, T., Hollemeyer, K., Schleheck, D. & Cook, A. M. Sulfoquinovose degraded by pure cultures of bacteria with release of C₃ organosulfonates: complete degradation in two-member communities. *FEMS Microbiol. Lett.* **328**, 39–45 (2012).
9. Cook, A. M. Biodegradation of s-triazine xenobiotics. *FEMS Microbiol. Rev.* **46**, 93–116 (1987).
10. Mayer, J. et al. 2,3-Dihydroxypropane-1-sulfonate degraded by *Cupriavidus pinatubonensis* JMP134: purification of dihydroxypropanesulfonate 3-dehydrogenase. *Microbiology* **156**, 1556–1564 (2010).
11. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
12. Graham, D. E., Xu, H. & White, R. H. Identification of coenzyme M biosynthetic phosphosulfolactate synthase: a new family of sulfonate biosynthesizing enzymes. *J. Biol. Chem.* **277**, 13421–13429 (2002).
13. Mampel, J. et al. A novel outer-membrane anion channel (porin) as part of the putatively two-component transport system for p-toluenesulfonate in *Comamonas testosteroni* T-2. *Biochem. J.* **383**, 91–99 (2004).
14. Mayer, J. & Cook, A. M. Homotaurine metabolized to 3-sulfofopropanoate in *Cupriavidus necator* H16: enzymes and genes in a patchwork pathway. *J. Bacteriol.* **191**, 6052–6058 (2009).
15. Saito, N. et al. Metabolite profiling reveals YihU as a novel hydroxybutyrate dehydrogenase for alternative succinic semialdehyde metabolism in *Escherichia coli*. *J. Biol. Chem.* **284**, 16442–16451 (2009).
16. Cornish-Bowden, A. Thermodynamic aspects of glycolysis. *Biochem. Educ.* **9**, 133–137 (1981).
17. Sugimoto, K., Sato, N. & Tsuzuki, M. Utilization of a chloroplast membrane sulfolipid as a major internal sulfur source for protein synthesis in the early phase of sulfur starvation in *Chlamydomonas reinhardtii*. *FEBS Lett.* **581**, 4519–4522 (2007).
18. Cook, A. M. & Denger, K. Dissimilation of the C₂ sulfonates. *Arch. Microbiol.* **179**, 1–6 (2002).
19. Benson, A. A. & Shibuya, I. Sulfocarbohydrate metabolism. *Fed. Proc.* **20**, 79 (1961).

Acknowledgements We thank E. Deuerling for substrain MG1655, J. Klebensberger for substrain BW25113 and its knockout mutants, and K. Leitner for help with growth experiments. The work of M.W. and A.-K.F. was supported by the Konstanz Research School Chemical Biology (KoRS-CB), the work of C.M. by German Research Foundation (DFG) grants (MA2436/4 and SFB766/A15) and by the Baden-Württemberg Stiftung (P-BWS-Glyko11), and the work of D.Sc. by a DFG grant (SCHL 1936/1-1) and by the University of Konstanz and the Konstanz Young Scholar Fund.

Author Contributions K.D. carried out most of the enzymic experiments, together with M.W. and A.-K.F., who carried out the heterologous expressions and RT-PCR. A.-K.F., A.S., C.M. and D.Sp. carried out the LC-MS analyses, and T.H. the chemical syntheses and NMR. D.Sc. set up the HILIC separation and carried out the 2D-PAGE, growth physiology and mutant analyses. A.M.C. and D.Sc. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.C. (alasdair.cook@uni-konstanz.de) or D.Sc. (david.schleheck@uni-konstanz.de).

Polarized release of T-cell–receptor–enriched microvesicles at the immunological synapse

Kaushik Choudhuri^{1*}, Jaime Llodrá^{2*}, Eric W. Roth³, Jones Tsai⁴, Susana Gordo^{5,6}, Kai W. Wucherpfennig^{5,6}, Lance C. Kam⁴, David L. Stokes^{2,7} & Michael L. Dustin^{8,9}

The recognition events that mediate adaptive cellular immunity and regulate antibody responses depend on intercellular contacts between T cells and antigen-presenting cells (APCs)¹. T-cell signalling is initiated at these contacts when surface-expressed T-cell receptors (TCRs) recognize peptide fragments (antigens) of pathogens bound to major histocompatibility complex molecules (pMHC) on APCs. This, along with engagement of adhesion receptors, leads to the formation of a specialized junction between T cells and APCs, known as the immunological synapse², which mediates efficient delivery of effector molecules and intercellular signals across the synaptic cleft³. T-cell recognition of pMHC and the adhesion ligand intercellular adhesion molecule-1 (ICAM-1) on supported planar bilayers recapitulates the domain organization of the immunological synapse^{4,5}, which is characterized by central accumulation of TCRs⁵, adjacent to a secretory domain², both surrounded by an adhesive ring^{4,5}. Although accumulation of TCRs at the immunological synapse centre correlates with T-cell function⁴, this domain is itself largely devoid of TCR signalling activity^{5,6}, and is characterized by an unexplained immobilization of TCR–pMHC complexes relative to the highly dynamic immunological synapse periphery^{4,5}. Here we show that centrally accumulated TCRs are located on the surface of extracellular microvesicles that bud at the immunological synapse centre. Tumour susceptibility gene 101 (TSG101)⁶ sorts TCRs for inclusion in microvesicles, whereas vacuolar protein sorting 4 (VPS4)^{7,8} mediates scission of microvesicles from the T-cell plasma membrane. The human immunodeficiency virus polyprotein Gag co-opts this process for budding of virus-like particles. B cells bearing cognate pMHC receive TCRs from T cells and initiate intracellular signals in response to isolated synaptic microvesicles. We conclude that the immunological synapse orchestrates TCR sorting and release in extracellular microvesicles. These microvesicles deliver transcellular signals across antigen-dependent synapses by engaging cognate pMHC on APCs.

The nature of the biophysical environment that governs molecular domain organization at the immunological synapse remains unclear. Confinement of pMHC⁴, TCRs⁵ and cytoplasm (Supplementary Fig. 1) suggests that a general diffusion barrier separates TCRs and cytoplasm at the immunological synapse centre from the rest of the T cell. To better understand the basis for the observed central confinement of pMHC, TCRs and cytoplasm at the immunological synapse, we investigated CD4⁺ T-cell immunological synapse formation using high-resolution optical imaging by total internal reflection fluorescence microscopy (TIRFM), integrated with transmission electron microscopy (TEM) and electron tomography.

T cells from TcrAND transgenic mice⁹ form TCR microclusters in response to engagement by the cognate class II pMHC molecule I-E^k

complexed with the moth cytochrome C peptide MCC 88–103 (MCC–I-E^k)⁶. Over approximately 10 min, TCR microclusters, together with bound pMHC^{4–6,9}, are transported on the cell surface to the immunological synapse centre, where they are consolidated into an immobilized domain⁶. To follow ultrastructural changes associated with immunological synapse formation, TcrAND T cells were fixed after 5, 10, 15 and 20 min of interaction with supported lipid bilayers containing MCC–I-E^k and ICAM-1, and imaged first by TIRFM and then by TEM. As a control, we used the non-cognate pMHC β_2m –I-E^k, which did not arrest motility or induce immunological synapse formation in TcrAND T cells (Supplementary Fig. 2a). TEM time series of TcrAND T cells forming immunological synapses on antigen-containing bilayers revealed changes in cell morphology that were characteristic of antigen-induced cell polarization (Supplementary Fig. 2b–d). Notably, at the 10 min time point, the centre of the T-cell contact interface showed an unexpected change in morphology, from a planar plasma membrane in continuous close apposition with the planar bilayer (Fig. 1a) to the appearance of numerous microvesicles (Fig. 1b and Supplementary Fig. 2e), approximately 70 nm in diameter (Supplementary Fig. 3a), that were contained within a central extracellular cavity (Fig. 1b). Microvesicle formation was antigen-specific, as they did not form with bilayers containing β_2m –I-E^k (Fig. 1a and Supplementary Fig. 2e), and could be modulated by the potency of the activating ligand, or by provision of costimulation¹⁰ (Fig. 1f and Supplementary Discussion). To visualize the distribution of microvesicles more clearly and verify their dissociation from the plasma membrane, we performed dual-axis tomography (Supplementary Video 1) on four serial sections through an immunological synapse, ranging from 150–250 nm in thickness. The associated three-dimensional model (Fig. 1c–e and Supplementary Videos 2 and 3) of the joined tomograms demonstrated that discrete extracellular microvesicles, with no connection to overlying plasma membrane (Supplementary Fig. 4a, b), predominate in the central cavity, along with occasional membrane projections and membrane buds of nascent microvesicles (Supplementary Fig. 4c–i).

Comparison of the distributions of TCRs and microvesicles at the immunological synapse demonstrated that they were spatially correlated (Supplementary Fig. 3b–d). To establish whether TCRs present at the immunological synapse centre were associated with microvesicles, we developed a novel method for optical–electron microscopy correlation, based on registration of TIRFM and corresponding TEM images of immunological synapses aligned using a microfabricated grid (Supplementary Fig. 5). Electron tomography of T-cell–bilayer interfaces confirmed the presence of microvesicles within a roughly circular extracellular cavity at the immunological synapse centre (Fig. 1h, i and Supplementary Videos 4 and 5). Optical–electron microscopy correlation

¹Program in Molecular Pathogenesis, Helen L. and Martin S. Kimmel Center for Biology and Medicine of the Skirball Institute of Biomolecular Medicine, 540 First Avenue, New York, New York 10016, USA.

²Program in Structural Biology, Helen L. and Martin S. Kimmel Center for Biology and Medicine of the Skirball Institute of Biomolecular Medicine, 540 First Avenue, New York, New York 10016, USA.

³Northwestern University Atomic and Nanoscale Characterization Experimental Center, Northwestern University, 2220 Campus Drive, Evanston, Illinois 60208, USA. ⁴Department of Biomedical Engineering, Columbia University, 500 W 120th Street, New York, New York 10027, USA. ⁵Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. ⁶Program in Immunology, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁷New York Structural Biology Center, 89 Convent Avenue, New York, New York 10027, USA. ⁸Department of Pathology, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA. ⁹Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, The University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7FY, UK.

*These authors contributed equally to this work.

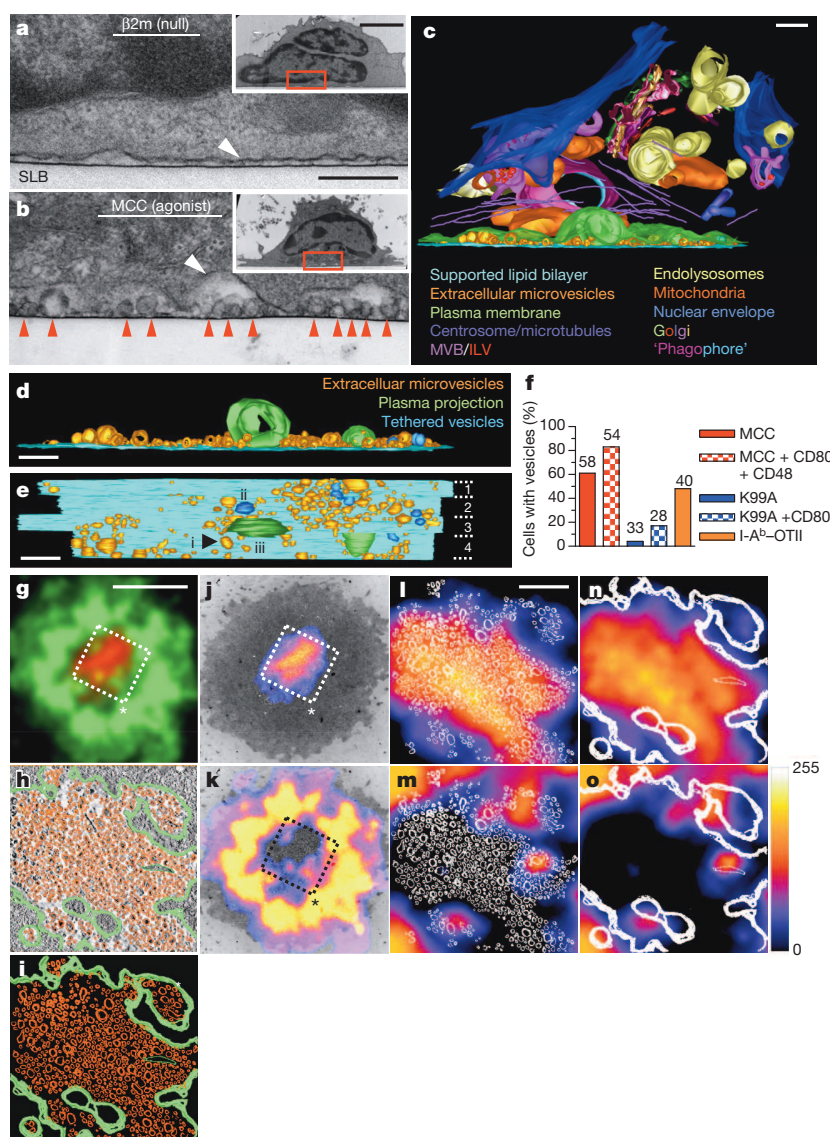


Figure 1 | Antigen-induced release of TCR-enriched microvesicles at the centre of immunological synapse. **a, b**, Transmission electron micrographs of TcrAND T cells interacting with supported lipid bilayers (SLB) containing ICAM-1, and the indicated pMHC. Inset, low magnification images of the same cells. Scale bar, 500 nm; inset scale bar, 2 μ m; red arrowheads, extracellular microvesicles; white arrowheads, plasma membrane. **c**, A three-dimensional ultrastructural model of the immunological synapse centre. The model was constructed from dual-axis tomograms of four serial sections made orthogonal to the bilayer from a single immunological synapse (numbered in **e**). ILV, intraluminal vesicle; MVB, multivesicular body; scale bar, 250 nm. **d**, Model from **c** with intracellular components and plasma membrane removed. Scale bar, 250 nm. **e**, Model from **d** rotated 90° in the *x* axis. Examples of an extracellular microvesicle (i), tethered vesicle (ii) and plasma membrane projection (iii) are indicated. Dotted lines, section boundaries; scale bar, 250 nm. **f**, Quantification of microvesicle production by TcrAND T cells in response to MCC-I-E^k (100 molecules per μ m²) or the weak agonist K99A-I-E^k (100 molecules per μ m²), or with additional co-stimulation provided by CD80 and CD48 on bilayers (200 molecules per μ m² for both). Microvesicle

production by OTII T cells, in response to ovalbumin-I-A^b pMHC ligand (100 molecules per μ m²), was also measured. *n* is indicated above bars. Results are pooled from two independent experiments. **g–o**, Optical-electron tomography correlation. **g**, Overlay of TCR (red) and F-actin (green) fluorescence in TIRFM images of the TcrAND T cell immunological synapse. Scale bar, 3 μ m, which also applies to **j** and **k**. **h**, A three-dimensional model of the immunological synapse centre constructed from an electron tomogram taken parallel to the T-cell–bilayer interface in **g**. T-cell plasma membrane, green; microvesicles, orange. The model is overlaid on a slice of the tomogram, and shown separately in **i** (scale bar in **i**, 500 nm, which also applies to **h**, **i** and **m–o**). **j, k**, Overlay of TCR (**j**) and F-actin (**k**) fluorescence from **g**, shown as a heat map, with the electron micrograph of the same T-cell–bilayer interface. Dashed boxes and asterisk in **g, j** and **k** indicate the orientation of the immunological synapse in relation to the tomogram and model overlays in **h, i** and **l–o**. **l, m**, Overlay of microvesicles modelled in **i** (white), scaled and aligned with the TCR fluorescence in **j** (**l**) and F-actin in **k** (**m**). **n, o**, Overlay of the T cell plasma membrane is modelled in **i** (white) scaled and aligned with the TCR fluorescence from **j** (**n**) and F-actin from **k** (**o**).

then allowed us to assign molecular distributions to immunological synapse ultrastructure. This revealed that TCR fluorescence at the immunological synapse (Fig. 1g, j) corresponded almost exclusively to extracellular microvesicles contained within the cavity at the immunological synapse centre (Fig. 1l, n). The characteristic F-actin depletion at the immunological synapse centre¹¹ (Fig. 1g, k) is also circumscribed by the plasma membrane boundary of the central cavity (Fig. 1o), reflecting interruption

of cytoplasmic continuity at the immunological synapse centre. The annular ICAM-1–LFA-1 adhesive domain, which surrounds centrally accumulated TCRs (and associated MCC–I-E^k), is wholly contained within the plasma membrane border of the central cavity (Supplementary Fig. 6), which effectively precludes membrane-tethered LFA-1 from binding ICAM-1 at the immunological synapse centre. Our finding that TCRs are present in extracellular microvesicles also accounts

for the inability of centrally accumulated TCRs to exchange with TCRs elsewhere in the T cell (Supplementary Fig. 1a, b), and for the confined cytoplasmic diffusion observed at the immunological synapse centre (Supplementary Fig. 1c).

After 30–45 min, TcrAND T cells break immunological synapse symmetry and resume motility¹². Migrating T cells left behind 80% of TCRs in particles on the bilayer (Supplementary Figs 7a, c and 8a and Supplementary Video 6). Optical–electron microscopy correlation of TcrAND T cells resuming motility after immunological synapse formation confirmed that these particles were extracellular TCR-enriched microvesicles released from the immunological synapse centre (Fig. 2a, b). The microvesicles diffused apart, which allowed us to characterize their composition by indirect immunofluorescence labelling and TIRFM (Supplementary Fig. 8b). TCR-enriched microvesicles were devoid of tyrosine phosphorylation (Fig. 2c), consistent with termination of TCR signalling. TCR-enriched microvesicles clustered MCC-I-E^k, but not β 2m-I-E^k, on bilayers (Fig. 2d and Supplementary Figs 7d and 9), establishing that TCRs were on the external surface of microvesicles and were capable of binding cognate pMHC.

To investigate the mechanism by which microvesicles were produced, we next studied the role of the endosomal sorting complex required for transport (ESCRT)-I member TSG101, which was previously implicated in ubiquitin-dependent sorting of TCRs to the immunological synapse centre⁶. ESCRT-I proteins, such as TSG101, typically carry out the first in a series of steps leading to vesicle budding and fission¹³. We suppressed TSG101 expression in TcrAND T cells with small interfering RNA (siRNA), and treated cells with non-targeting small oligo-RNA (sRNA) as a control (Supplementary Fig. 10a). As expected, TcrAND T

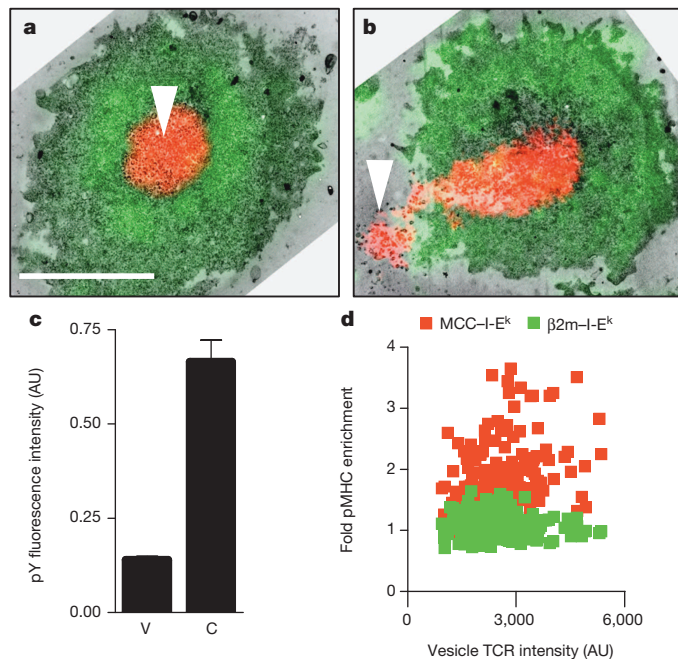


Figure 2 | TCR-enriched microvesicles are post-signalling extracellular products of T-cell activation that retain pMHC-binding competence. **a, b**, Optical–electron microscopy correlation of sessile TcrAND T cells forming an immunological synapse in response to MCC-I-E^k and ICAM-1 (**a**) and resuming motility (**b**), showing the distributions of F-actin (green) and TCRs (red) in relation to released microvesicles deposited on the bilayer (arrowheads). Scale bar, 4 μ m. **c**, Quantification of phosphotyrosine (pY) fluorescence intensity in arbitrary units (AU) of released TCR-enriched microvesicles (V) and their originating cells (C) detected by indirect immunofluorescence labelling and TIRFM. **d**, Enrichment of fluorescently labelled cognate MCC-I-E^k or non-binding β 2m-I-E^k by released TCR-enriched microvesicles. Fold pMHC enrichment is defined as: (pMHC fluorescence intensity colocalized with TCR-enriched microvesicles)/(free bilayer pMHC fluorescence intensity). Error bars represent s.e.m.

cells, in which TSG101 expression was suppressed, failed to translocate TCRs, together with bound MCC-I-E^k, to the immunological synapse centre, whereas central TCR accumulation was unaffected by treatment with control sRNA (Supplementary Fig. 10b, c). Optical–electron microscopy correlation revealed that suppression of TSG101 resulted in accumulation of TCRs, together with bound MCC-I-E^k, in a ring around the central cavity defining the immunological synapse centre (Fig. 3a), whereas treatment with control sRNA had no effect on TCR incorporation into microvesicles (Supplementary Fig. 10d). Suppression of TSG101 reduced, but did not abolish, microvesicle production (Supplementary Fig. 11). Notably, the remaining microvesicles present at the immunological synapse centre were devoid of TCRs, demonstrating that TSG101 was necessary for sorting of TCRs into microvesicles (Fig. 3a).

We next investigated the role of the ESCRT-III complex, which is responsible for the final steps in vesicle fission. We used primary human

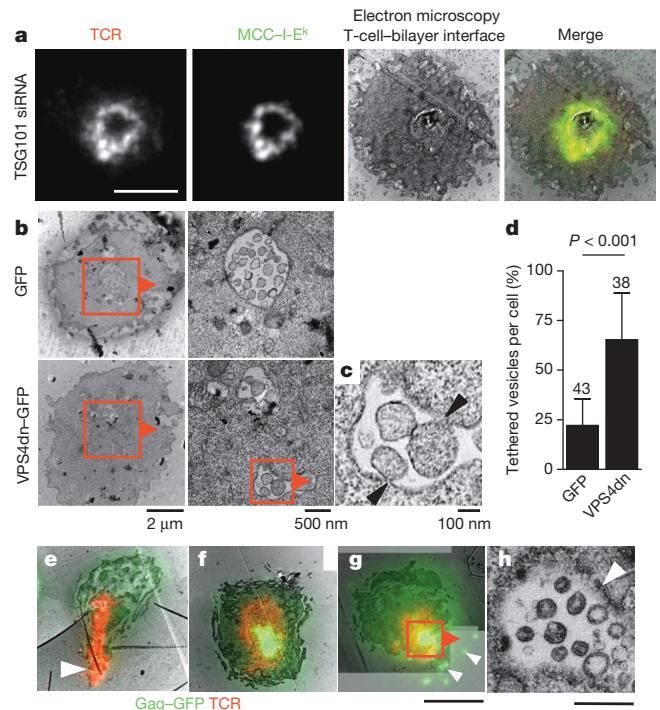
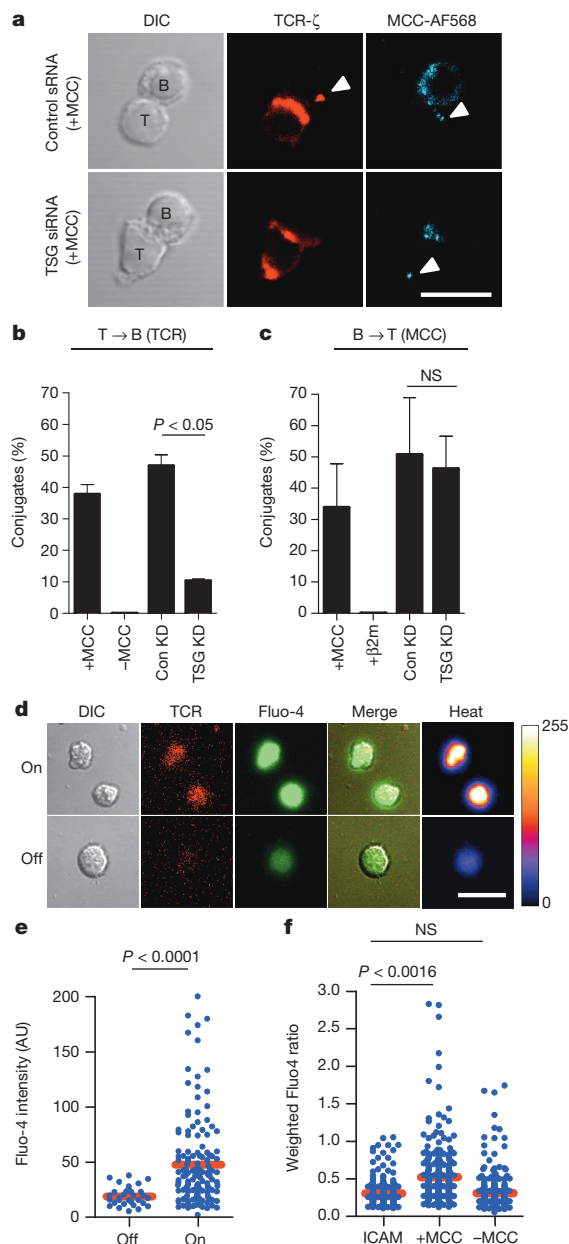


Figure 3 | Biogenesis of TCR-enriched microvesicles is mediated by ESCRT proteins and antagonized by HIV Gag. **a**, Optical–electron microscopy correlation of TCR and MCC-I-E^k distributions of a representative TcrAND T cell following siRNA-mediated TSG101 suppression. Scale bar, 3 μ m. **b**, Electron micrographs of the T-cell-bilayer interface of human CD4⁺ T cells forming an immunological synapse in response to TCR engagement. Cells were transfected with a construct encoding dominant-negative VPS4 fused to GFP (VPS4dn-GFP) or GFP only. Red boxes and arrowheads indicate regions imaged at higher magnification in panels on the right. Arrowheads in **c** indicate nascent vesicles tethered to the plasma membrane. **d**, Quantification of limiting membrane-tethered microvesicles expressed as a percentage of total microvesicles within central compartments of the immunological synapse in cells expressing indicated constructs. Means and s.d. are shown, *P* value is for Student's *t*-test. *n* is indicated above data bars. Results are pooled from two independent experiments. **e–g**, Representative optical–electron microscopy correlation of the immunological synapse in human CD4⁺ T cells transfected with a construct encoding HIV Gag fused to GFP (Gag-GFP, green). Red, TCR; scale bar, 3 μ m. **e**, T cell without central Gag-GFP accumulation resuming migration and releasing TCR-enriched microvesicles (arrowhead). **f**, T cell forming an immunological synapse with centrally accumulated Gag-GFP. **g**, T cell with centrally accumulated Gag-GFP resuming motility and releasing Gag-GFP-containing microvesicles (arrowheads). **h**, Higher magnification image of boxed region in **g** showing internal juxta-membrane density in Gag-containing microvesicles. Arrowhead, plasma membrane; scale bar, 500 nm.



CD4⁺ T cells which, like TcrAND murine T cells, produce TCR-enriched microvesicles at the immunological synapse centre (Supplementary Fig. 12). We transfected human CD4⁺ T cells with a construct encoding a dominant-negative mutant of human VPS4(E228Q), that prevents ATP binding, fused to green fluorescent protein (VPS4dn-GFP)¹³. Expression of VPS4dn-GFP disrupts the function of endogenous VPS4¹³, which binds to and catalytically disassembles¹⁴ membrane-associated ESCRT-III oligomers and is thought to be obligatory for sustaining membrane budding reactions^{15,16}. As a control, we transfected human CD4⁺ T cells with a construct encoding GFP alone. Control T cells formed microvesicles at the immunological synapse centre, the majority of which were separated from the limiting plasma membrane of the central cavity (Fig. 3b, d). In marked contrast, the immunological synapse of VPS4dn-GFP-transfected T cells contained nascent microvesicles, within a fragmented central compartment, which remained tethered to the limiting plasma membrane (Fig. 3b–d) at constricted bud necks (Fig. 3c). Consistent with the role of VPS4 'downstream' of early acting ESCRTs, VPS4dn-GFP did not affect TCR sorting to the immunological synapse centre (Supplementary Fig. 13), although its distribution at the immunological synapse was altered when compared to wild-type VPS4 (Supplementary

Figure 4 | TSG101 selectively controls TCR transfer to B cells that signal in response to pMHC engagement by microvesicle-tethered TCRs. **a**, Confocal microscopy of TcrAND T cells, treated with TSG101 suppressing siRNA (TSG siRNA) or control oligo-RNA (Control siRNA), forming conjugates with congenic splenic B cells loaded with fluorescent MCC peptide (cyan). $T = 30$ min.; red, TCR- ζ ; arrowheads, transferred TCR and fluorescent MCC peptide; scale bar, 10 μ m. **b**, Quantification of TCR transfer to B cells in T-B conjugates. **c**, Quantification of MCC peptide transfer to T cells in T-B conjugates. NS, not significant. Means from three independent experiments are shown in **b**, **c**. Error bars represent s.d. β 2m, non-stimulating control peptide. Con KD, control siRNA; TSG KD, knockdown with TSG101 siRNA. **d**, Live B cells loaded with calcium-sensitive fluorescent dye Fluo-4 were imaged by confocal microscopy on bilayers containing deposited TCR-enriched microvesicles ($T = 30$ min). Panels show differential interference contrast (DIC) and fluorescence images of TCR-enriched microvesicle patches on bilayers (red) and Fluo-4 fluorescence in B cells (green and heat map). Shown are examples of B cells interacting with TCR-enriched microvesicles (On), or on bilayer areas without detectable TCR-enriched microvesicles (Off). Scale bar, 10 μ m. **e**, Quantification of Fluo-4 fluorescence intensity of B cells 'On' or 'Off' TCR-enriched microvesicles. Data are representative of two experiments. **f**, Quantification of Fluo-4 intensity of B cells pulsed with MCC peptide or without antigen, after 30 min incubation on bilayers containing ICAM-1 alone, or with bilayers containing deposited TCR-enriched microvesicles. Fluorescence values were divided by the mean Fluo-4 intensity of B cells treated with 1 μ M ionomycin at the end of the experiment. Data pooled from two independent experiments. Blue dots, fluorescence intensity in individual cells; red bars, sample mean; P value, one-way analysis of variance corrected for all comparisons.

Fig. 14). These observations indicate that VPS4 is involved in fission of TCR-rich membrane buds, thereby producing TCR-enriched microvesicles at the immunological synapse centre^{13,17}.

The ESCRT pathway is also exploited by the human immunodeficiency virus (HIV) structural polyprotein Gag^{8,13} for membrane budding and release of virus particles from the plasma membrane of infected cells. This prompted us to test whether Gag might interfere with the biogenesis of TCR-enriched microvesicles in human CD4⁺ T cells. We expressed a construct encoding Gag fused to GFP (Gag-GFP) in primary human CD4⁺ T cells, and analysed immunological synapse formation by optical-electron microscopy correlation. Transfected T cells adhered poorly to supported bilayers containing ICAM-1 alone, with small Gag-GFP puncta present throughout the contact interface (Supplementary Fig. 15b, c). Notably, TCR engagement led to robust recruitment of Gag-GFP to the contact interface (Supplementary Fig. 15b, c), where it accumulated at the immunological synapse centre, displaced TCRs (Fig. 3f and Supplementary Fig. 15d, e), and resulted in the release of Gag-GFP-containing microvesicles, which are analogous to virus-like particles (VLP) that are produced during HIV infection (Fig. 3g, h and Supplementary Fig. 15f, g). Conversely, in T cells with low Gag-GFP expression, TCRs accumulated normally at the immunological synapse centre (Supplementary Fig. 15d, e), resulting in the release of TCR-enriched microvesicles (Fig. 3e and Supplementary Fig. 15f, g). The ESCRT-binding domain of Gag is required for its recruitment and central accumulation at the immunological synapse following TCR engagement (Supplementary Fig. 16 and Supplementary Discussion). Taken together, these findings indicate that Gag antagonizes ESCRT-dependent sorting of TCRs into microvesicles, and in this context TCR ligation directs the polarized budding of VLPs at the centre of the immunological synapse.

To verify that microvesicles have a physiological role in cellular interactions, we next asked whether TcrAND T cells produce TCR-enriched microvesicles in conjugates with antigen-bearing B cells. As expected, TcrAND T cells readily accumulated TCRs at the immunological synapse centre in conjugates with MCC-pulsed congenic B10.Br B cells (Supplementary Fig. 17a, b). In approximately 40% of T-B cell conjugates (Fig. 4a, b), small puncta of TCRs could be detected in B cells 'distal' to the immunological synapse, consistent with T cell to B cell synaptic transfer of TCR-containing microvesicles, whereas no T-cell polarization or TCR transfer was detected in conjugated B cells pulsed with

β2m peptide (Supplementary Fig. 17a, b). Because TSG101 is critical for the production of TCR-enriched microvesicles in the planar bilayer model, we next investigated whether TCR transfer to live B cells is affected by siRNA-mediated suppression of TSG101⁶. In conjugates of TSG101-suppressed TcrAND T cells and MCC-pulsed B cells, transfer of TCRs was reduced by approximately 80% (Fig. 4a, b), relative to conjugates treated with control siRNA, despite effective TCR polarization to the immunological synapse (Supplementary Fig. 17b). As previously described, transfer of pMHC from B cells to T cells was also observed (Fig. 4a, c and Supplementary Fig. 17a)^{18,19}. However, this pMHC transfer was unaffected by TSG101 silencing in the T cell (Fig. 4a, c), demonstrating that TSG101 selectively controls TCR transfer to B cells during bidirectional membrane exchange in T–B cell conjugates¹⁹.

We next asked whether isolated microvesicles induce signalling in antigen-presenting B cells. Engagement of MHC II molecules on B cells triggers tyrosine phosphorylation and intracellular calcium signalling, by coupling to the B-cell antigen receptor signalling machinery²⁰, resulting in cell proliferation in primed B cells²¹. We therefore asked whether B cells presenting cognate MCC–I-E^k complexes are activated by specific TCRs present on microvesicles. We introduced B cells, loaded with the Ca²⁺-sensitive dye Fluo-4, onto bilayers containing TCR-enriched microvesicles, from which T cells had been removed (Supplementary Fig. 17c), and monitored Ca²⁺ signalling by confocal microscopy (Fig. 4d and Supplementary Fig. 17d). B cells pulsed with MCC were motile on bilayers, but stopped and showed sustained increase in intracellular Ca²⁺ on encountering TCR-enriched microvesicle patches on bilayers (Fig. 4d, e). In contrast, B cells not loaded with MCC did not show an increase in Ca²⁺ relative to controls (that is, cells on bilayers with ICAM-1 alone, Fig. 4d, f). Thus, B-cell signalling by pMHC is initiated by cognate recognition of TCRs on microvesicles. In support of a role for TCR-enriched microvesicles during human T–B cell interaction, we found that Raji B cells, in superantigen-induced conjugates with human CD4⁺ T cells²², received TCR puncta that activate phospholipase Cγ1 (PLCγ1), a key mediator of intracellular Ca²⁺ signalling²³ (Supplementary Fig. 18).

Our finding that the immunological synapse centre is an extracellular cavity, filled with TCR-enriched microvesicles by an ESCRT-dependent mechanism, provides a fundamentally new model for supramolecular domain organization at the immunological synapse (Supplementary Fig. 19a). Shedding of TCRs in microvesicles constitutes a novel mechanism for TCR ‘downregulation’, following engagement by pMHC, that acts in parallel with receptor internalization²⁴. Our observations raise the possibility that other immune cells known to accumulate immunoreceptors at the synapse centre, such as B²⁵ and NK²⁶ cells, may also release them in microvesicles for intercellular communication. Cognate recognition of TCR-enriched microvesicles by pMHC on B cells may provide ‘help’ to B cells that is calibrated to the pMHC density present on their surface, pointing to a plausible mechanism for avidity-adjusted delivery of T-cell help to B cells²⁷. Finally, we find that TCR-enriched microvesicle biogenesis is a native pathway in T cells, triggered by antigen recognition, that may be co-opted by HIV proteins for polarized retroviral transmission at antigen-dependent immunological synapses (Supplementary Fig. 19b).

METHODS SUMMARY

Murine TcrAND T cells were expanded with MCC peptide and IL-2, and used once quiescent. Human CD4⁺ T cells were isolated from peripheral blood by negative selection and used within 48 h. T cells clones were expanded for one cycle using heterologous irradiated peripheral blood mononuclear cells (PBMNCs), phytohaemagglutinin (PHA) and IL-2 and used once quiescent. 1,2 dioleoyl-*sn*-glycero-3-phosphatidylcholine (DOPC) lipid bilayers containing 12.5 mol% 1,2-dioleoyl-*sn*-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt) (DGS-NTA-Ni) were deposited on coverslips that were cleaned with peroxidized sulphuric acid and rinsed with pure water. Bilayers contained 100 molecules per μm² MCC–I-E^k, and 200 molecules per μm² ICAM-1 unless otherwise stated. Optical imaging was performed using a Nikon TIRF microscope and an LSM510 confocal microscope. TEM imaging and tomography was performed using Phillips CM12,

CM200 and FEI Technai F20 microscopes. For optical–electron microscopy correlation, lipid bilayers were formed on coverslips containing chrome registration grids on which cells were deposited and fixed for imaging. Following TIRFM imaging of the entire grid, samples were embedded in Epon, and sectioned parallel to the bilayer. TEM images of the first parallel section contained the chrome grid imprint, allowing location, imaging and alignment of electron micrographs and tomograms with TIRFM images. Cells were transfected with DNA or RNA by nucleofection (Amaxa).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 August 2012; accepted 12 December 2013.

Published online 2 February 2014.

1. Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* **314**, 537–539 (1985).
2. Monks, C. R. F., Freiberg, B. A., Kupfer, H., Sciaky, N. & Kupfer, A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**, 82–86 (1998).
3. Stinchcombe, J. C., Majorovits, E., Bossi, G., Fuller, S. & Griffiths, G. M. Centrosome polarization delivers secretory granules to the immunological synapse. *Nature* **443**, 462–465 (2006).
4. Grakoui, A. et al. The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).
5. Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M. L. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. *Immunity* **25**, 117–127 (2006).
6. Vardhana, S., Choudhuri, K., Varma, R. & Dustin, M. L. Essential role of ubiquitin and TSG101 protein in formation and function of the central supramolecular activation cluster. *Immunity* **32**, 531–540 (2010).
7. Babst, M., Wendland, B., Estepa, E. J. & Emr, S. D. The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J.* **17**, 2982–2993 (1998).
8. von Schwedler, U. K. et al. The protein network of HIV budding. *Cell* **114**, 701–713 (2003).
9. Kaye, J. et al. Selective development of CD4⁺ T cells in transgenic mice expressing a class II MHC-restricted antigen receptor. *Nature* **341**, 746–749 (1989).
10. Linsley, P. S. et al. Binding of the B cell activation antigen B7 to CD28 costimulates T cell proliferation and interleukin 2 mRNA accumulation. *J. Exp. Med.* **173**, 721–730 (1991).
11. Kaizuka, Y., Douglass, A. D., Varma, R., Dustin, M. L. & Vale, R. D. Mechanisms for segregating T cell receptor and adhesion molecules during immunological synapse formation in Jurkat T cells. *Proc. Natl Acad. Sci. USA* **104**, 20296–20301 (2007).
12. Sims, T. N. et al. Opposing effects of PKCθ and WASp on symmetry breaking and relocation of the immunological synapse. *Cell* **129**, 773–785 (2007).
13. Garrus, J. E. et al. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**, 55–65 (2001).
14. Sakseena, S., Wahlman, J., Teis, D., Johnson, A. E. & Emr, S. D. Functional reconstitution of ESCRT-III assembly and disassembly. *Cell* **136**, 97–109 (2009).
15. Wollert, T., Wunder, C., Lippincott-Schwartz, J. & Hurley, J. H. Membrane scission by the ESCRT-III complex. *Nature* **458**, 172–177 (2009).
16. Stuchell-Brereton, M. D. et al. ESCRT-III recognition by VPS4 ATPases. *Nature* **449**, 740–744 (2007).
17. Sachse, M., Strous, G. J. & Klumperman, J. ATPase-deficient hVPS4 impairs formation of internal endosomal vesicles and stabilizes bilayered clathrin coats on endosomal vacuoles. *J. Cell Sci.* **117**, 1699–1708 (2004).
18. Martínez-Martin, N. et al. T cell receptor internalization from the immunological synapse is mediated by TC21 and RhoG GTPase-dependent phagocytosis. *Immunity* **35**, 208–222 (2011).
19. He, T. et al. Bidirectional membrane molecule transfer between dendritic and T cells. *Biochem. Biophys. Res. Commun.* **359**, 202–208 (2007).
20. Lang, P. et al. TCR-induced transmembrane signaling by peptide/MHC class II via associated Ig-α/β dimers. *Science* **291**, 1537–1540 (2001).
21. Cambier, J. C. & Lehmann, K. R. Ia-mediated signal transduction leads to proliferation of primed B lymphocytes. *J. Exp. Med.* **170**, 877–886 (1989).
22. Mittelbrunn, M. et al. Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nature Commun.* **2**, 282 (2011).
23. Marshall, A. J., Niino, H., Yun, T. J. & Clark, E. A. Regulation of B-cell activation and differentiation by the phosphatidylinositol 3-kinase and phospholipase Cγ pathway. *Immunol. Rev.* **176**, 30–46 (2000).
24. Liu, H., Rhodes, M., Wiest, D. L. & Vignali, D. A. On the dynamics of TCR:CD3 complex cell surface expression and downmodulation. *Immunity* **13**, 665–675 (2000).
25. Fleire, S. J. et al. B cell ligand discrimination through a spreading and contraction response. *Science* **312**, 738–741 (2006).
26. Vanherberghen, B. et al. Human and murine inhibitory natural killer cell receptors transfer from natural killer cells to target cells. *Proc. Natl Acad. Sci. USA* **101**, 16873–16878 (2004).

27. Victora, G. D. *et al.* Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* **143**, 592–605 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Schütz for suggesting the photoactivation experiment, H. Chen for microfabrication, W. Sundquist for providing Gag–GFP, VPS4–GFP and VPS4dn–GFP constructs, P. Bieniasz for Gag–mCherry, GagΔL–mCherry, ALIX–GFP and CHMP4B–GFP constructs, the New York Structural Biology Center for electron microscopy tomography, support and instrumentation and members of the Dustin laboratory for helpful discussions and contributions of reagents. We thank J. Nance for the gift of biotinylated duramycin-linked biotin. This work was supported in part by a Cancer Research Institute fellowship and NIH grant K99AI093884 (K.C.), a Wellcome Trust Principal Research Fellowship (M.L.D.), a Kennedy Trust Senior

Research Fellowship (M.L.D.) and NIH grants AI043542, AI045757, AI055037, AI088377, AI093884 and EY016586 (Nanomedicine Development Center).

Author Contributions K.C., M.L.D. and D.L.S. conceived the study, K.C. and J.L. designed and performed the experiments, E.W.R. performed sectioning, L.C.K. and J.T. designed and fabricated optical–electron microscopy reference grids, K.W.W. and S.G. made essential reagents, K.C. and M.L.D. wrote the manuscript. All authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.L.D. (michael.dustin@kennedy.ox.ac.uk) or D.L.S. (stokes@nyu.edu).

SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation

Lianna M. Johnson^{1*}, Jiamu Du^{2*}, Christopher J. Hale¹, Sylvain Bischof¹, Suhua Feng^{1,3}, Ramakrishna K. Chodavarapu¹, Xuehua Zhong^{1†}, Giuseppe Marson², Matteo Pellegrini¹, David J. Segal⁴, Dinshaw J. Patel² & Steven E. Jacobsen^{1,3}

RNA-directed DNA methylation in *Arabidopsis thaliana* depends on the upstream synthesis of 24-nucleotide small interfering RNAs (siRNAs) by RNA POLYMERASE IV (Pol IV)^{1,2} and downstream synthesis of non-coding transcripts by Pol V. Pol V transcripts are thought to interact with siRNAs which then recruit DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) to methylate DNA^{3–7}. The SU(VAR)3-9 homologues SUVH2 and SUVH9 act in this downstream step but the mechanism of their action is unknown^{8,9}. Here we show that genome-wide Pol V association with chromatin redundantly requires SUVH2 and SUVH9. Although SUVH2 and SUVH9 resemble histone methyltransferases, a crystal structure reveals that SUVH9 lacks a peptide-substrate binding cleft and lacks a properly formed S-adenosyl methionine (SAM)-binding pocket necessary for normal catalysis, consistent with a lack of methyltransferase activity for these proteins⁸. SUVH2 and SUVH9 both contain SRA (SET- and RING-ASSOCIATED) domains capable of binding methylated DNA⁸, suggesting that they function to recruit Pol V through DNA methylation. Consistent with this model, mutation of DNA METHYLTRANSFERASE 1 (MET1) causes loss of DNA methylation, a nearly complete loss of Pol V at its normal locations, and redistribution of Pol V to sites that become hypermethylated. Furthermore, tethering SUVH2 with a zinc finger to an unmethylated site is sufficient to recruit Pol V and establish DNA methylation and gene silencing. These results indicate that Pol V is recruited to DNA methylation through the methyl-DNA binding SUVH2 and SUVH9 proteins, and our mechanistic findings suggest a means for selectively targeting regions of plant genomes for epigenetic silencing.

To gain insights into the function of SUVH2/SUVH9, we solved the crystal structure of an amino-terminally truncated SUVH9 construct (residues 134–650), which contains all the known functional domains (the SRA, pre-SET and SET domains) (Fig. 1a, Extended Data 1a and Supplementary Table 1). The structure of SUVH9 is composed of three segments: a two-helix bundle towards the N terminus (residues 138–194), the SRA domain (residues 195–379) and the pre-SET/SET domains (residues 380–637). There are extensive inter-domain interactions that can stabilize the overall architecture of the protein (Fig. 1a and Extended Data Fig. 1b–g).

The SRA domain of SUVH9 resembles those of UHRF1 and SUVH5^{10–13} (Fig. 1b). Based on the SUVH5 SRA–mCHH (mCHH, methylated CHH, where H refers to A, T or C) DNA complex structure¹³, we modelled a mCHH DNA into SUVH9 (Extended Data Fig. 2a). The DNA could be positioned in the nucleic-acid-binding cleft of the SRA domain without significant steric clashes and the proposed flipped-out 5-methylcytosine (5mC) base readily inserts into the binding pocket of the SRA domain.

Although SUVH9 contains histone methyltransferase pre-SET and SET domains similar to Dim5, G9a and GLP^{14–16} (Extended Data Fig. 2b), it shows neither detectable histone methyltransferase activity nor binding capacity for the SAM cofactor *in vitro*⁸. SUVH9 and SUVH2 lack

post-SET domains that are normally critical for cofactor- and peptide-substrate-binding, as well as catalysis (Extended Data Fig. 3). Compared with the structure of human H3K9 methyltransferase GLP (Fig. 1c, d), the putative S-adenosylhomocysteine (SAH)-binding pocket and peptide-binding channel of SUVH9 are very open and incompletely formed (Fig. 1e, f) and cannot stably retain either a bound SAH molecule or the peptide substrate, especially in the absence of the stabilizing role of the post-SET domain. In summary, the SUVH9 structure is consistent with the demonstrated ability of SUVH9 to bind to methylated DNA, and supports the view that SUVH9 (and SUVH2) proteins encode inactive methyltransferase homologues.

We sought to further characterize the role of SUVH2 and SUVH9 in the RNA-directed DNA methylation (RdDM) pathway by determining their effect on siRNAs genome-wide. Most siRNAs are eliminated in Pol IV (*nprp1*) mutants, whereas only some siRNAs are reduced in Pol V (*nprp1*) mutants¹⁷ (Fig. 2a). We found that the *suvh2 suvh9* double mutant reduced siRNA abundance at siRNA clusters that were dependent on both Pol IV and Pol V, but not at Pol V-independent clusters (Fig. 2a, b). Thus, the *suvh2 suvh9* double mutant siRNA phenotype closely resembles that of a Pol V mutant. We next sought to determine whether SUVH2/SUVH9 might be involved in the production of non-coding transcripts by Pol V. At two characterized sites (IGN22 and P6)^{18,19} we found that *suvh2 suvh9* reduced Pol V transcripts to the same extent as *nprp1* (Fig. 2c). Similar results were observed at the *AtSN1* locus in *suvh2 suvh9*⁹. To test whether SUVH2/SUVH9 are required for Pol V chromatin occupancy, we used chromatin immunoprecipitation (ChIP) of a Flag-tagged NRPE1 (the largest subunit of Pol V). We observed only background levels of Pol V binding at IGN5 and IGN22 in *suvh2 suvh9* compared to a sixfold enrichment in wild type (WT; Fig. 2d). We further analysed our ChIPs by next generation sequencing (ChIP-seq) and found that binding of Pol V at all previously identified sites¹⁸ was significantly decreased or eliminated in *suvh2 suvh9* (Fig. 2e, f).

To determine the effect of SUVH2/SUVH9 on DNA methylation at defined Pol V binding sites, we used whole-genome bisulphite sequencing (BS-seq). As in *nprp1*, CHH methylation at Pol V binding sites was eliminated in *suvh2 suvh9* (Fig. 3a). We also analysed BS-seq of the single mutants *suvh2* and *suvh9* to determine whether SUVH2/SUVH9 act redundantly at all sites or have non-overlapping sites where they function. We found that *suvh2* had a stronger effect than *suvh9* at Pol V sites as well as at differentially methylated regions (DMRs) defined in either *suvh2* or *suvh9* single mutants, or in the *suvh2 suvh9* double mutant (Extended Data Fig. 4a–c). These results indicate that SUVH2 and SUVH9 act redundantly throughout the genome to control RdDM.

Our results suggest that a reinforcing loop exists between DNA methylation and Pol V binding via SUVH2/SUVH9. To further test this model we used a mutation in the maintenance methyltransferase MET1 that eliminates CG methylation genome-wide and also reduces CHG and

¹Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, California 90095, USA. ²Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, California 90095, USA. ⁴Genome Center and Department of Biochemistry and Molecular Medicine, University of California at Davis, Davis, California 95616, USA. [†]Present address: Wisconsin Institute for Discovery, Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA.

*These authors contributed equally to this work.

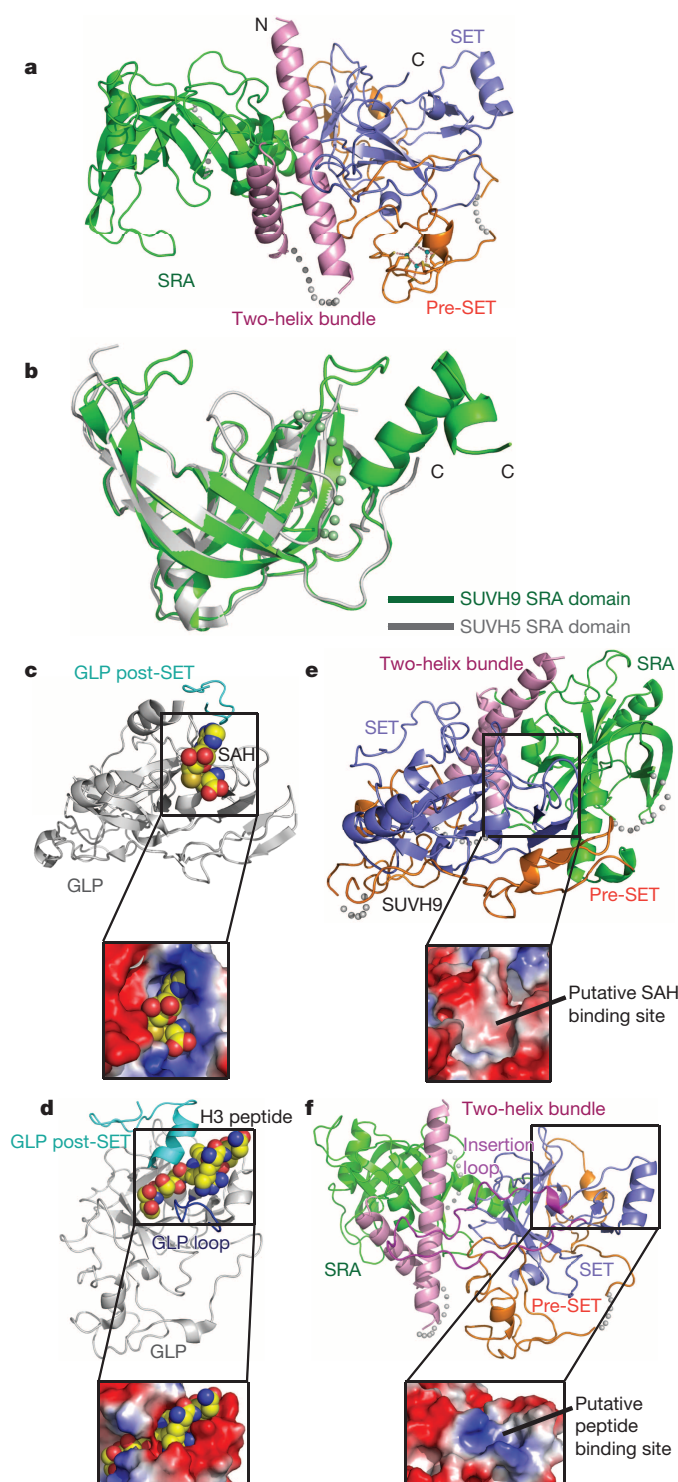


Figure 1 | Crystal structure of SUVH9. **a**, Ribbon diagram of the SUVH9 crystal structure containing a two-helix bundle, SRA domain, pre-SET domain and SET domain coloured in pink, green, orange and blue, respectively. The Zn_3Cys_9 cluster is highlighted in a ball-and-stick representation and disordered regions are shown with dashed lines. **b**, A superposition of SUVH9 SRA domain (in green) and SUVH5 SRA domain (in silver) shows that both proteins adopt a similar fold. **c**, Top, crystal structure of human GLP in complex with bound SAH (PDB code 2IGQ) in a silver ribbon representation. Bottom, the SAH binding site in an electrostatic surface representation. The cofactor SAH is shown in a space-filling representation. **d**, Top, the crystal structure of human GLP in complex with SAH and H3K9me2 peptide (PDB code 2RF1) in silver ribbon representation. Bottom, the peptide binding site in an electrostatic surface representation. The post-SET domain and the acidic loop of the SET domain involved in peptide-substrate binding are highlighted in cyan and dark blue, respectively. The bound peptide is shown in a space-filling representation in both panels. **e**, Top, the crystal structure of SUVH9 in the free state in a colour-coded ribbon representation. Bottom, an expanded view of the putative SAH binding site in an electrostatic surface representation. **f**, Top, the crystal structure of SUVH9 in the free state in a colour-coded ribbon representation. Bottom, an expanded view of the putative peptide-binding site in an electrostatic surface representation. The long insertion loop of the SET domain is highlighted in magenta.

To test directly whether SUVH2/SUVH9 may be sufficient to recruit Pol V, we used a zinc finger (ZF) to tether SUVH2 to an unmethylated epiallele of *FWA*, *fwa-4*. *FWA* is normally silenced owing to DNA methylation of tandem repeats in its promoter²³. *FWA* epialleles have lost this methylation, leading to ectopic expression of *FWA* and a heritable late flowering phenotype^{23,24}. Surprisingly, siRNAs are still observed in *fwa* epialleles, but are unable to direct DNA methylation²⁵. We proposed that Pol V transcripts may be missing in *fwa* epialleles and that recruitment of Pol V by SUVH2 might therefore stimulate *FWA* methylation and silencing. To test this we transformed a ZF-SUVH2 fusion protein construct into the *fwa-4* epiallele (Extended Data Fig. 10a). As negative controls, we also transformed a haemagglutinin (HA)-tagged SUVH2 line without the ZF (HA-SUVH2) as well as a construct in which the ZF was fused to KRYPTONITE (KYP/SUVH4) (a SUVH protein not required for RdDM).

Approximately 75% of the *fwa-4* plants transformed with the ZF-SUVH2 (T1 generation) flowered early as compared to the parental *fwa-4* line, suggesting silencing of *FWA* (Fig. 4a). The *fwa-4* line transformed with HA-SUVH2 or ZF-KYP flowered at the same time as the *fwa-4* parent, showing that the effect was specific to the ZF-SUVH2 fusion. Flowering time was measured in the T2 generation confirming these observations (Fig. 4b). The presence of the control ZF-KYP at *FWA* was shown by ChIP (Extended Data Fig. 10b); however, we were unable to detect the ZF-SUVH2, most likely owing to its instability or low abundance.

We next used bisulphite sequencing to determine whether *FWA* gene silencing was associated with DNA methylation. In wild type, DNA methylation was detected over a large region, whereas in both *fwa-4* and transformants with ZF-KYP or HA-SUVH2, this region was devoid of DNA methylation (see ZF-KYP; Fig. 4c). In three independent ZF-SUVH2 T1 lines, DNA methylation was observed immediately around the Zn finger binding sites in all cytosine sequence contexts (Extended Data Fig. 10c). We analysed one of the lines (ZF-SUVH2-2) in the T2 and T3 generations using BS-seq and found that methylation extended approximately 150 base pairs in either direction from the binding sites and did not expand significantly between generations (Fig. 4c). *FWA* methylation and gene silencing were maintained in T2 plants that had segregated away the ZF-SUVH2 transgene (Extended Data Fig. 10d), indicating that targeting by SUVH2 is capable of inducing DNA methylation and gene silencing that can be maintained in the absence of the initial trigger.

To determine whether Pol V was present at *FWA*, we used NRPE1 ChIP. As expected, we found enrichment of Pol V at two known Pol V sites, IGN5 and IGN22, in the wild-type and *fwa-4* lines, but not in *nrpe1* mutant plants (Fig. 4d). At *FWA*, we found enrichment of Pol V

CHH methylation^{20–22} (Fig. 3b). Using endogenous antibodies to NRPE1, ChIP-seq revealed that Pol V occupancy was virtually eliminated in *met1* at sites normally occupied by Pol V (Fig. 3c and Extended Data Figs 5–7). By contrast, at sites previously identified as gaining methylation in *met1*²¹, we observed an increase in Pol V binding (Fig. 3d, e and Extended Data Figs 8 and 9). Furthermore, point mutations in the SRA domains of SUVH2 and SUVH9 were shown to cause a loss of RdDM⁸. Together, these results indicate that SUVH2/SUVH9 binding to methylated DNA recruits Pol V, thus providing a link between pre-existing DNA methylation and the recruitment of further methylation by RdDM.

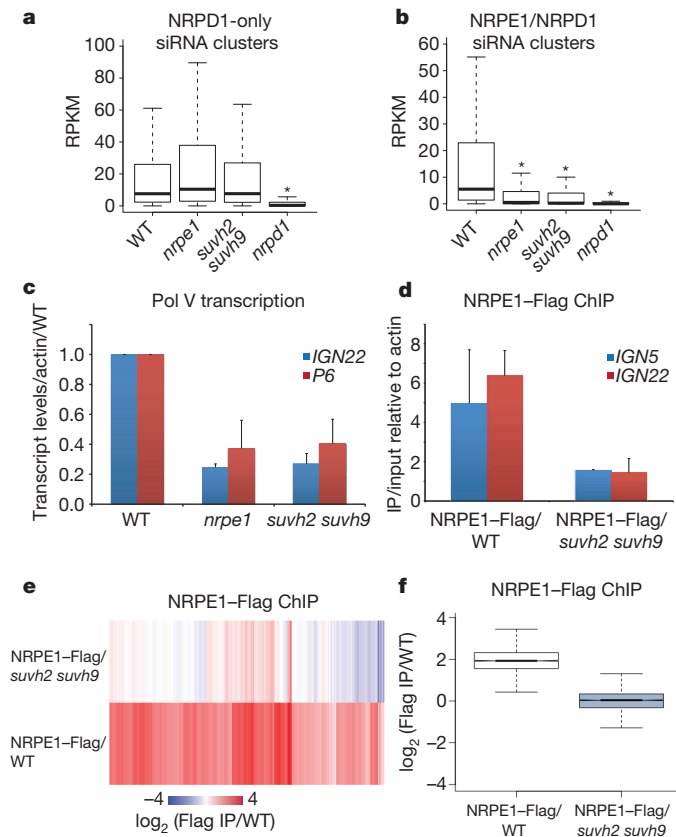


Figure 2 | SUVH2 and SUVH9 are required for Pol V-dependent siRNA production, chromatin binding and transcription. **a**, Boxplot (whiskers extend to ± 1.5 interquartile range (IQR)) of reads per kilobase per million reads (RPKM) values for 24-nucleotide siRNAs at previously defined siRNA clusters dependent on Pol IV (NRPD1), but not Pol V (NRPE1). 24-nt counts allow for up to 100 identical reads to be counted at any given position. * indicates a significant decrease ($P < 2.2 \times 10^{-16}$, Mann-Whitney *U*-test). **b**, Similar to plot in **a** for clusters defined as dependent on Pol IV and Pol V. **c**, Quantitative PCR with reverse transcription (RT-qPCR) of *IGN22* and *P6* relative to *ACTIN7* and normalized to Columbia-0 (WT). Mean \pm standard deviation (s.d.) of two biological replicates. **d**, Quantitative PCR (qPCR) of *IGN22* and *IGN5* from Flag ChIP shown as enrichment of IP/input relative to *ACTIN7* in NRPE1-Flag/WT and NRPE1-Flag/*svuh2 svuh9* lines. Mean \pm s.d. of two biological replicates. **e**, Heat map of NRPE1 enrichment at defined NRPE1 sites determined by Flag ChIP-seq in either NRPE1-Flag/WT or NRPE1-Flag/*svuh2 svuh9*, with Flag ChIP in WT as negative control. **f**, Box plot of NRPE1 enrichment at sites shown in Fig. 2e for NRPE1-Flag/WT and NRPE1-Flag/*svuh2 svuh9*.

in the wild type at both the promoter and transcript regions, but not in *nrpe1* or *fwa-4* (Fig. 4d). However, in the ZF-SUVH2-transformed *fwa-4* plants we could now see enrichment of Pol V at *FWA* (Fig. 4d), indicating that SUVH2 is sufficient to recruit Pol V.

To look for a direct interaction between SUVH2/SUVH9 and Pol V, we queried several IP-mass spectrometric data sets from purifications of NRPE1-Flag, but failed to find any SUVH2/SUVH9 peptides. However, we did identify SUVH2 peptides in two independent mass spectrometric data sets from DRD1 purifications⁶ (Supplementary Table 2). DRD1 is a component of the DDR complex (also containing DMS3 and RDM1) which interacts with Pol V⁶ and is required for Pol V occupancy throughout the genome¹⁸. The number of SUVH2 peptides observed was lower than those from the DMS3 and RDM1 proteins and also lower than the level of peptides of most Pol V complex components, indicating that the interaction between SUVH2 and DRD1 is weaker or more transient than the interaction between the DDR components or between DDR and Pol V. To confirm the interaction, we performed co-immunoprecipitation

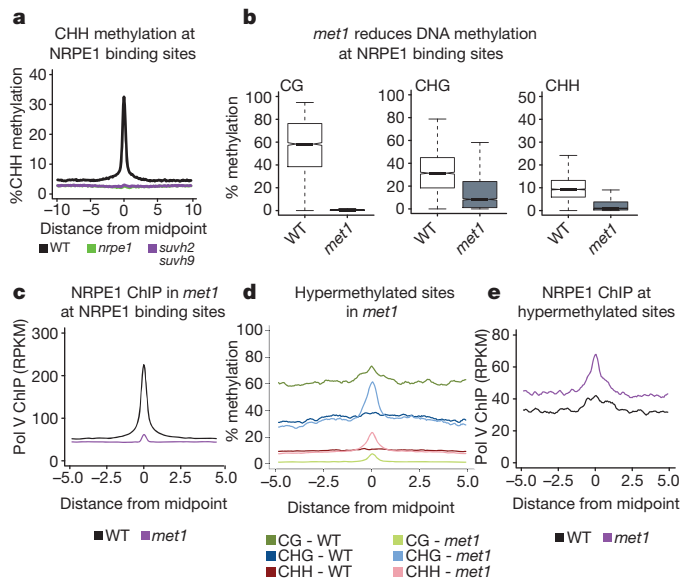


Figure 3 | Pol V binding is dependent on DNA methylation. **a**, Metaplot of per cent CHH methylation at all defined NRPE1 binding sites as determined by BS-seq in wild type (WT), *nrpe1* and *svuh2 svuh9*. **b**, Box plots showing DNA methylation in each cytosine context at defined NRPE1 binding sites in WT and *met1*. **c**, Pol V occupancy in *met1* is reduced at NRPE1 sites. **d**, Metaplot of DNA methylation at sites defined as hyper methylated in *met1*. **e**, Pol V occupancy in *met1* is increased at defined hyper methylated sites.

experiments with HA-tagged SUVH2 and Flag-tagged DRD1 in transgenic *Arabidopsis* plants and were able to detect HA-SUVH2 in a Flag-DRD1 pull-down (Fig. 4e). We were also able to detect Flag-DRD1 in a pull-down using HA-SUVH2 expressed in leaves of *Nicotiana benthamiana* and purified on HA magnetic beads (Extended Data 10e). These results confirm the IP-mass spectrometry observations and are consistent with a model in which SUVH2 acts indirectly via a transient interaction with DRD1 to recruit Pol V.

Because RdDM is triggered at genomic locations that synthesize both Pol IV-dependent siRNAs and Pol V-dependent non-coding transcripts, it is critical to understand the mechanisms that direct Pol IV and Pol V to chromatin. Furthermore, because RdDM functions to silence transposons that do not necessarily share consensus DNA sequences, a strong prediction is that Pol IV and Pol V will use epigenetic information in their targeting. We propose that SUVH2/SUVH9 serve as critical components of this targeting system by recruiting Pol V to DNA methylation. We recently showed that recruitment of Pol IV involves SHH1, a protein that binds to histone H3 lysine 9 methylation, a silencing mark also found at RdDM sites^{26–28}. Therefore, the dual marks of DNA methylation and histone H3 lysine 9 methylation seem to act in self-reinforcing loop mechanisms to maintain Pol V and Pol IV at sites of RdDM. These mechanistic findings also suggest a means for selectively targeting DNA methylation by directing both siRNAs and Pol V to overlapping regions of genome, paving the way to chromatin engineering in crop plants.

Similar to other components of RdDM, SUVH2/SUVH9 are required for the establishment of *de novo* DNA methylation during transformation of an unmethylated *FWA* gene into *Arabidopsis*⁸. An interesting question is how an originally unmethylated DNA is first recognized as an RdDM target if SUVH2/SUVH9 DNA methylation-mediated Pol V recruitment cannot initially occur. One possibility is that shortly after the transgene is integrated into the genome, a permissive chromatin environment may allow RNA Pol V (or another polymerase) to produce surveillance transcripts that might serve to initiate RdDM through siRNA-mediated recruitment of DRM2. After this initial round of DNA methylation, SUVH2/SUVH9 could then bind methylated DNA

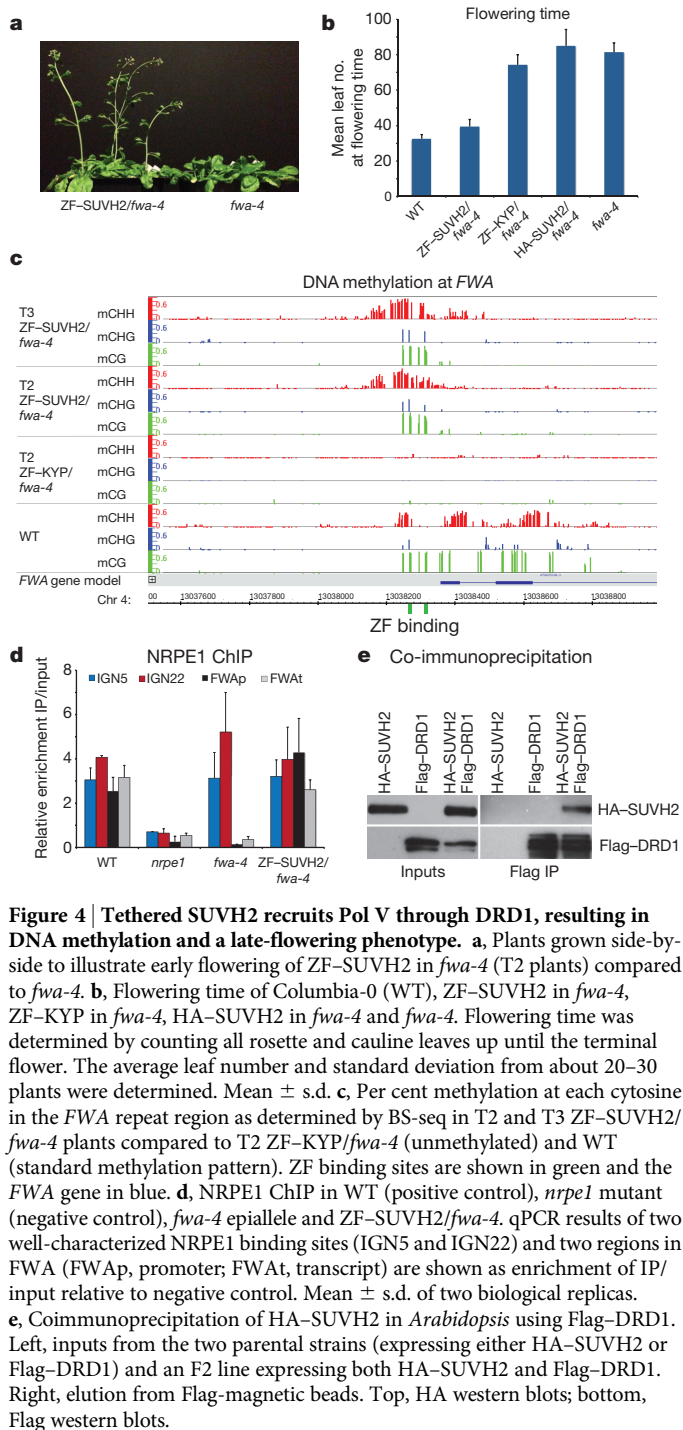


Figure 4 | Tethered SUVH2 recruits Pol V through DRD1, resulting in DNA methylation and a late-flowering phenotype. **a**, Plants grown side-by-side to illustrate early flowering of *ZF-SUVH2* in *fwa-4* (T2 plants) compared to *fwa-4*. **b**, Flowering time of Columbia-0 (WT), *ZF-SUVH2* in *fwa-4*, *ZF-KYP* in *fwa-4*, *HA-SUVH2* in *fwa-4* and *fwa-4*. Flowering time was determined by counting all rosette and cauline leaves up until the terminal flower. The average leaf number and standard deviation from about 20–30 plants were determined. Mean \pm s.d. **c**, Per cent methylation at each cytosine in the *FWA* repeat region as determined by BS-seq in T2 and T3 *ZF-SUVH2/fwa-4* plants compared to T2 *ZF-KYP/fwa-4* (unmethylated) and WT (standard methylation pattern). ZF binding sites are shown in green and the *FWA* gene in blue. **d**, NRPE1 ChIP in WT (positive control), *nrpe1* mutant (negative control), *fwa-4* epiallele and *ZF-SUVH2/fwa-4*. qPCR results of two well-characterized NRPE1 binding sites (IGN5 and IGN22) and two regions in *FWA* (FWAp, promoter; FWAt, transcript) are shown as enrichment of IP/input relative to negative control. Mean \pm s.d. of two biological replicates. **e**, Coimmunoprecipitation of HA-SUVH2 in *Arabidopsis* using Flag-DRD1. Left, inputs from the two parental strains (expressing either HA-SUVH2 or Flag-DRD1) and an F2 line expressing both HA-SUVH2 and Flag-DRD1. Right, elution from Flag-magnetic beads. Top, HA western blots; bottom, Flag western blots.

and recruit Pol V for subsequent rounds of transcription. In this way, a self-reinforcing loop would be established in which pre-existing DNA methylation is required for the maintenance of RNA-directed DNA methylation.

METHODS SUMMARY

All biological materials and extended methods are described in detail in the Methods. siRNA libraries were made from flowers of the indicated lines as described previously²⁶ and Pol V transcription assays were performed as described previously¹⁸. ChIP-seq, BS-seq, co-immunoprecipitation, protein purification and crystallization, and data analysis are described in the Methods. Supplementary Table 1 provides crystallography data, Supplementary Table 2 provides IP-mass spectrometry data and Supplementary Table 3 provides primers used in this study.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 May; accepted 2 December 2013.

Published online 22 January 2014.

- Péllissier, T. & Wassenaar, M. A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. *RNA* **6**, 55–65 (2000).
- Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, A. J. & Matzke, M. RNA-directed DNA methylation in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99** (Suppl. 4), 16499–16506 (2002).
- Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* **11**, 204–220 (2010).
- Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
- Pikaard, C. S., Haag, J. R., Ream, T. & Wierzbicki, A. T. Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci.* **13**, 390–397 (2008).
- Law, J. A. *et al.* A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in *Arabidopsis*. *Curr. Biol.* **20**, 951–956 (2010).
- Wierzbicki, A. T., Ream, T. S., Haag, J. R. & Pikaard, C. S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genet.* **41**, 630–634 (2009).
- Johnson, L. M., Law, J. A., Khattar, A., Henderson, I. R. & Jacobsen, S. E. SRA-domain proteins required for DRM2-mediated de novo DNA methylation. *PLoS Genet.* **4**, e1000280 (2008).
- Kuhlmann, M. & Mette, M. F. Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in *Arabidopsis thaliana*. *Plant Mol. Biol.* **79**, 623–633 (2012).
- Hashimoto, H. *et al.* The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455**, 826–829 (2008).
- Avvakumov, G. V. *et al.* Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455**, 822–825 (2008).
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**, 818–821 (2008).
- Rajakumara, E. *et al.* A dual flip-out mechanism for 5mC recognition by the *Arabidopsis* SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes Dev.* **25**, 137–152 (2011).
- Zhang, X. *et al.* Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**, 117–127 (2002).
- Zhang, X. *et al.* Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell* **12**, 177–185 (2003).
- Wu, H. *et al.* Structural biology of human H3K9 methyltransferases. *PLoS ONE* **5**, e8570 (2010).
- Mosher, R. A., Schwach, F., Studholme, D. & Baulcombe, D. C. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl Acad. Sci. USA* **105**, 3145–3150 (2008).
- Zhong, X. *et al.* DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nature Struct. Mol. Biol.* **19**, 870–875 (2012).
- Wierzbicki, A. T., Haag, J. R. & Pikaard, C. S. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**, 635–648 (2008).
- Aufsatz, W., Mette, M. F., Matzke, A. J. & Matzke, M. The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. *Plant Mol. Biol.* **54**, 793–804 (2004).
- Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).
- Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
- Soppe, W. J. *et al.* The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791–802 (2000).
- Kakutani, T. Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in *Arabidopsis thaliana*. *Plant J.* **12**, 1447–1451 (1997).
- Chan, S. W., Zhang, X., Bernatavichute, Y. V. & Jacobsen, S. E. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol.* **4**, e363 (2006).
- Law, J. A. *et al.* Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385–389 (2013).
- Zhang, H. *et al.* DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc. Natl Acad. Sci. USA* **110**, 8290–8295 (2013).
- Law, J. A., Vashisht, A. A., Wohlschlegel, J. A. & Jacobsen, S. E. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet.* **7**, e1002195 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to W. Shi at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL) for support in diffraction data collection. We thank C. Pikaard for the NRPE1 antibodies and M. Akhavan for technical assistance. High-throughput sequencing was performed in the UCLA BSCRC

BioSequencing Core Facility. This work was supported by the Abby Rockefeller Mauze Trust and the Maloris and STARR foundations to D.J.P., and NIH grant GM60398 to S.E.J. C.J.H. is supported by the Damon Runyon post-doctoral fellowship, S.B. is supported by a post-doctoral fellowship of the Swiss National Science Foundation, S.F. is a Special Fellow of the Leukemia & Lymphoma Society, and X.Z. is supported by Ruth L. Kirschstein National Research Service grant F32GM096483-01. S.E.J. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions Protein purification and crystallography were designed and performed by J.D., G.M. and D.J.P., siRNA libraries were made by S.B., whole-genome bisulphite sequencing was performed and analysed by S.F., L.M.J., C.H., R.K.C. and M.P., Pol V transcription assays were performed by X.Z., ChIP-seq was performed and

analysed by L.M.J. and C.H., cloning and flowering time experiments were performed by L.M.J., the Zn finger was designed by D.J.S. and S.E.J., pull-down experiments were performed by L.M.J. and S.B., the manuscript was written by L.M.J., J.D., D.J.P. and S.E.J., and S.E.J. participated in all experimental design.

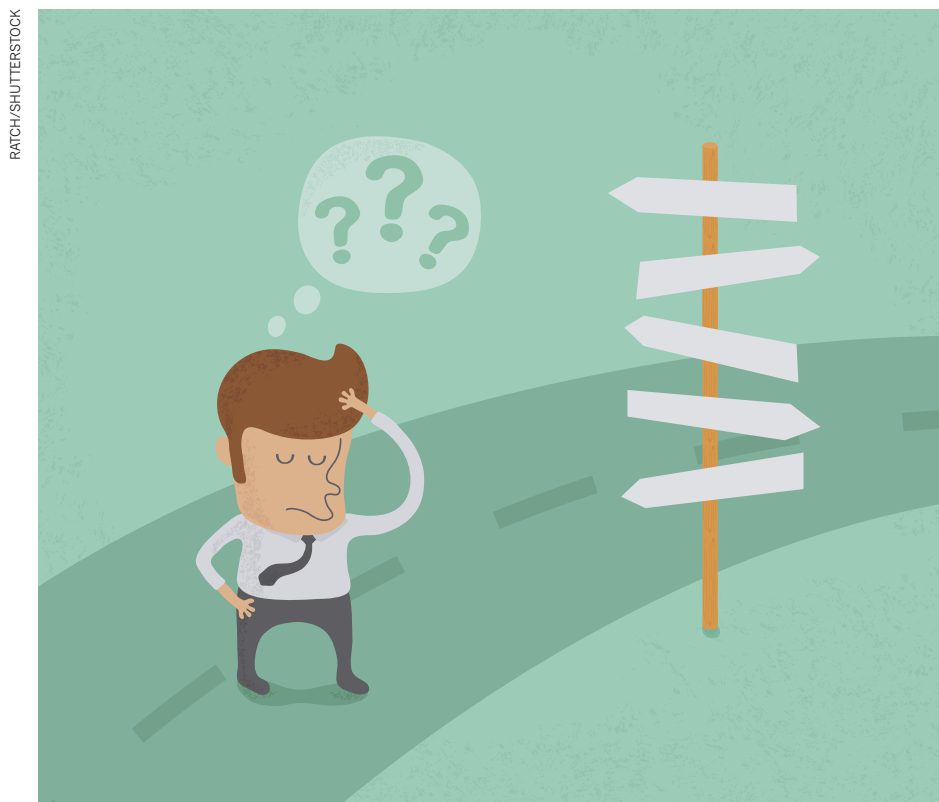
Author Information The coordinates and structure factors of SUVH9 have been deposited in the RCSB Protein Data Bank with the accession code 4NJ5. Sequencing data have been deposited at GEO (GSE52041). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.E.J. (jacobsen@ucla.edu) or D.J.P. (pateld@mskcc.org).

CAREERS

COLUMN The key ways to engage an audience p.131

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com



CHINA

At a crossroads

China is luring back expatriates with generous incentives, causing many to weigh the pros and cons of returning.

BY QUIRIN SCHIERMEIER

A year ago, Jun Yao was facing a difficult dilemma. Having completed a PhD and two postdoctoral stints in the United States, the Chinese-born neuroscientist was well placed to pursue an academic career in his adopted country. But recruiters in China had also taken note of his achievements and were offering him substantial incentives to continue his research back home. Should he stay or go?

When Yao had left China in 2002, the United States had obvious advantages as a place for launching his career. He did a PhD at Pennsylvania State University (Penn State) in College

Park, followed by postdocs at the University of Wisconsin-Madison and at the Salk Institute for Biological Studies in La Jolla, California. But he knew that the environment for science research, and especially funding, had notably improved in his homeland over the past ten years. So much so, in fact, that Tsinghua University in Beijing — one of the nation's top-tier institutions — was now able to offer him a position as principal investigator in neuroscience and a generous starting package. Based on what Yao had seen of offers in the United States, the deal was exceptional.

Yao mulled over the best way forward, and in the end he opted to return. (Soon after, a US university expressed strong interest. But Yao

declined, having made his decision.) During interviews for the job, he had been impressed by the high calibre and hard-working nature of the students. "I had not made up my mind before I went to Beijing for the interviews," he says. "But when I saw the new labs and modern facilities they have built, and when I spoke with students and scientists at Tsinghua, I decided that this is the place for me to do science in the future."

China is counting on Western-trained researchers such as Yao to help the nation boost its scientific status to a level similar to that of the United States, Europe and Japan. And it is forking out the money to do so: in 2012, the nation spent 1.98% of its gross domestic product on research. That is slightly higher than Europe's allocation for research and, with the amount growing by approximately 20% a year, China's proportional spending on research edges ever closer to the 2.79% of the United States and the 3.34% of Japan.

But China's political environment sometimes gives pause to expatriates who are accustomed to living and working in a democratic society — the nation's human-rights and freedom-of-speech issues worry some. Also of concern are the quality of science at some of China's lesser-known institutes and universities; the level of expectation that returnees often face as a foreign-trained expert; and the fact that funding streams still often depend on connections to 'old-boy networks'. "For Chinese scientists who have been abroad for years, it is difficult to be reintegrated into the system," says Cong Cao, a sociologist and expert on Chinese science who is based at the University of Nottingham, UK.

TALENT CONTEST

To bring in the talent, China is embracing science and coming up with concrete plans to lure back promising expats. Its 'Thousand Talents' plan — to which Yao successfully applied with assistance from the university — aims to entice foreign-trained professionals to dedicate their skills to building the government's vision for a twenty-first-century China, which, according to statements from vice-president Li Yuanchao, includes both wealth creation and technological exchange with other nations. As part of a special section of the Thousand Talents plan, scientists, scholars and engineers under 40 years old who have at least three years of postdoctoral experience at a reputable non-Chinese university can apply for a position at a Chinese university or industry laboratory. Successful applicants get grant money ►

► and relocation funds worth several million renminbi (depending on the research project); research projects typically run for three years. Candidates apply to the university or institute of their choice, and successful applicants then negotiate a salary and terms of employment with their host institute. Attracted by the prospect of lucrative funding and new career opportunities, about 3,000 scientists have heeded the call since the scheme was established in 2008.

By all indications, there will be plenty of applicants for years to come. China's scientific diaspora numbers well over 400,000 scientists and other scholars. In the United States, for example, there is scarcely a research department that does not have Chinese-born PhD students and postdocs within its ranks. Many Chinese researchers leave their home country in the hope of finding permanent jobs — in or outside academia — and many do. But as funding opportunities in the United States and elsewhere decline, they are increasingly considering career opportunities back home.

"I've been in the United States for 20 years now and never has it been more difficult to secure grant money, especially for early-career scientists," says Yao's former PhD supervisor Gong Chen, a neurobiologist at Penn State in University Park. "I do tell my students that it's a good time to do science in China — but I also tell them about the downsides."

PROCEED WITH CAUTION

Those downsides are not inconsiderable, and require careful study of any potential new workplace. "Go to China," recommends Yao, "and look around before you apply for a position or sign a contract."

For instance, department heads and research administrators often have unrealistic expectations about what repatriates can accomplish in a short time, says Chen. "There is immense pressure on returnees to instantly produce scientific results and publish in prestigious journals," he says. "One is right to expect hard work and determination from young scientists. But they need more time and support than they are often being granted."

Newcomers will typically need to start their research from scratch in more-or-less empty laboratories and with little knowledge of local circumstances. "Be social — good relationships with colleagues and superiors are essential, after all," says Chen. "But do also prepare for difficulties and delays that you will inevitably face in a new environment, no matter how hard you are willing to work."

One of the main concerns is where the position is based. Only a few Chinese institutions — among them Tsinghua University and Shanghai's Fudan and Jiao Tong universities — offer conditions that are comparable to those at top Western institutes. Researchers there can count on state-of-the-art lab equipment, top-quality collaborations with foreign institutions, adherence to the precepts of research integrity,



Expatriates might find it difficult to reintegrate into the Chinese system, says sociologist Cong Cao.

the absence of old-boy-network practices and a high calibre of students, faculty and visitors. But at less-esteemed universities in smaller cities, the level of science is often second-rate, says David Zweig, chair of social sciences at the Hong Kong University of Science and Technology. He suggests avoiding universities that allow professors to keep on their own PhD students routinely as they become postdocs and professors. "You only want to go to places that send students out to the world."

Returnees should also seek out a mentor who is familiar with local circumstances and can help them to acclimatize and to deal with problems if they arise. "As an outsider, you'd otherwise be pretty much lost," Zweig says. "You might not even be able to get any research money."

He and others say that at many institutes, getting funded or promoted often depends on forming personal relationships with local research administrators or party officials. Reports of plagiarism, theft of ideas and other fraudulent behaviour are common. All of this can add up to a situation in which Chinese scientists are more productive when they are abroad than when they are in China, says Rao Yi, a biologist formerly with Northwestern University's medical school in Chicago, Illinois, who became dean of the School of Life Sciences at Peking University in Beijing in 2007. "If they care about their research and integrity, those who might return should go for institutions that have reformed or are determined to reform," he says.

Generous benefits to returnees often cause *hong yan bing* — 'red-eye disease', as envy is often called in China — among other Chinese academics. The Thousand Talents plan makes no specific provision as to the terms of employment of returnees, such as salary or period of employment; those terms differ depending on the host institution. But high-profile awardees can get instant tenure and a salary that is considerably higher than the 200,000 renminbi (US\$32,650) per year of an average Chinese

professor. Such efforts, says Yi, sometimes cause Chinese professors to view senior returnees as competitors for funds and fame. Cao agrees: "There are often tensions between people with and without foreign education," he says. "Frankly, some do not want expats to return to China permanently."

China expert Denis Simon, vice-provost of the Office of International Strategic Initiatives at Arizona State University in Phoenix, says that the problem is especially difficult when the recruit is not Chinese-born. Another government incentive — the 'Recruitment Program of Foreign Experts' — encourages scientists with a PhD from an overseas university to work in China for at least nine months a year over a minimum of three consecutive years, and it offers rewards that are similar to those of the Thousand Talents scheme. "It's one thing to attract Chinese-born postgraduates from overseas," he says. "Bringing senior scientists to China is a different story. It's hard to swallow for China's old guard that Mr Smith gets so much when their Mr Wang gets so much less."

For his part, Yao knows that challenges lie ahead, but he has found much to encourage him. He has been assigned a 180-square-metre laboratory on the Tsinghua campus on the outskirts of Beijing — more research space than he ever had in the United States. It will take a good few months to buy the equipment and hire the technicians and postdocs he needs. To remain busy and productive, he maintains an active research alliance with the group he left behind in California.

PREPARING TO RETURN

Chinese researchers would be wise to establish a scientific reputation for themselves before they consider going back to their homeland, says Cao. "Don't move too early in your career," he says. "Make sure you have done at least a few years of productive postdoctoral research and produced some nice papers, else you'll have the air of one who is not competitive in the West."

The Thousand Talents plan also includes a popular short-term option for scientists who are reluctant to sever ties with the West altogether. The awardees who select this route keep their overseas academic jobs and obligations while working up to three months per year — usually during the summer break — at a Chinese host institute.

Those moving back should consider the scientific opportunities and family-friendly provisions of, for example, the Thousand Talents plan, which often includes premium medical service, free insurance and eligibility to purchase property at discount rates. But they should also be aware that a successful application is no guarantee of a permanent job in science. The tenure track to a permanent academic position — so familiar in the United States — is widely absent in China. In the past, says Chen, pretty much anyone who obtained a PhD could look forward to a job for life. But leading universities in Beijing and Shanghai have begun to evaluate junior faculty members' performance after five years and to make promotions to permanent positions on the basis of those evaluations.

The Chinese government — always struggling to reconcile communist ideals with the transition to a market economy — is aware of the grumbling within its



"I do tell my students that it's a good time to do science in China — but I also tell them about the downsides."

Gong Chen

academic institutions. But the Communist Party has made it clear that bringing back the best is crucial to the nation's economic development. During his 2007–12 term as head of the Communist Party's organization department, Li Yuanchao stressed that bringing back expatriates was a "strategic investment".

Critics say, however, that if China really wants

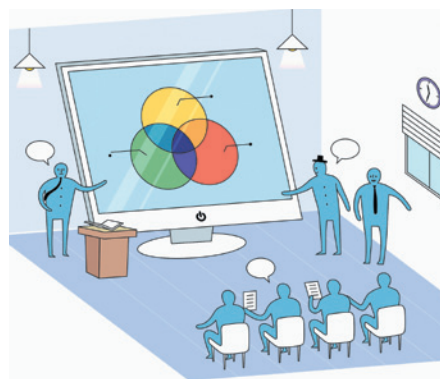
to bring back the best, it must overhaul its science culture and its academic institutions to embrace democratic pluralism and improve transparency. The idea is that more returnees from the West would not just help the development of Chinese science, but that of society as well. "The inflow of young science talent from abroad," says Chen, "will help shape our nation." ■

Quirin Schiermeier is Nature's Germany correspondent.

COLUMN

A powerful narrative

Scientists should find engaging ways to present information to their target audience, says **Yoshimi Rii**.



Last summer, I took part in an unfortunate episode on the first day of a teacher workshop. Along with three other scientists, I gave talks to schoolteachers on the use of real-time scientific data in the classroom. Because we didn't want to talk down to our audience, we kept to the standard scientific format for our presentations, with minor adjustments to specific jargon. As the day went on, however, it became apparent that we had missed the mark.

"Why do scientists' presentations all look the same?" said a teacher to me during the break. "It's like you guys all get the same template." Her words struck me like a 10-tonne anvil. She was right. Where was the creativity?

Most of us at one time will have attended a talk about what should have been an engaging topic, only to find ourselves being lulled to sleep. Most likely, the speaker was using PowerPoint or Keynote. There are better approaches.

MAKING THE DISCONNECTION

The user-friendly nature of software such as PowerPoint allows anyone to make a presentation without much thought. Type in a few talking points, throw in some pictures and voilà! Here's my cookie-cutter talk. Never mind that there's too much information on some slides — they're bulleted, so they must be easy to understand! The plot makes five or six points, but hey, it's colourful and impressive, so that's OK, right? And let's not worry about those bumpy transitions, never mind the lack of a coherent narrative.

Reliance on bullet points and complicated graphs has caused many to become lax at applying important performance skills. Disconnected from the public, scientists often forget to explain what to them is everyday logic. Last September,

at a press conference at the University of Hawaii in Manoa for the Intergovernmental Panel on Climate Change Working Group I Report, scientists showed modelled rises in sea level from now until 2100. I watched the eyes of the reporters roll backwards in their sockets.

Because slides are easy to recycle from one presentation to another, researchers often give a talk on autopilot using slides that they've shown hundreds of times before. If the presenter is disengaged, how can they expect the audience to listen?

BREAKING BAD

On the second day of the teacher workshop, we decided to scrap our prepared talks and start afresh. I placed a quote from *Moby Dick* in the title box and was rewarded with a smile from an English teacher in the audience. I then tossed a microbe-shaped stuffed toy to a teacher in the front row to open up a discussion about bacteria and phytoplankton. We asked them how they wanted to illustrate these concepts to their students. For the remainder of the workshop, the talks became less structured, more interactive and better appreciated. I vowed never to rely on cookie-cutter presentations again.

Last October, I found myself at a foundation symposium with an audience of retired doctors, professors and other distinguished society members who had made donations in support of research at the university. I sat next to one donor, a 90-year-old veteran who was part of the highly decorated 442nd Regimental Combat Team in the Second World War. As part of the symposium, senior scientists and I were invited to give talks on exciting research being conducted at our university. I was nervous.

But I found myself breathing a sigh of relief. In the first talk, a scientist studying the history of calendars opened up a star-studded umbrella. Another scientist presented a beautiful slide show of mushroom pictures and riveted the audience with accounts of his hunts in unusual places. At the end of the symposium, the veteran next to me grabbed my hands and said, "I feel great knowing that my money has made all of this possible. Thank you." We relayed our message that day, and without the aid of a single graph. ■

Yoshimi Rii is a graduate student in phytoplankton ecology at the University of Hawaii in Manoa.

TONGRO/TONGRO IMAGES/CORBIS

UNIT SIMULATION

Reconnect with your past.

BY COY ST CLAIR

The street was empty for as far as Jane could see, and she was resigned to another night without a customer. This would be three nights in a row. Yuri would not be happy. Jane pulled her coat closer against the evening chill and began walking.

By the time she reached the end of her territory, three cars had passed without slowing down, even though she had looked hopefully at each one. It was insulting, really. She was a professional, after all. She may have been dressed in normal clothes, but there was no mistaking what she was, especially in this part of town. As she turned to make a slow walk down to the other end of her territory, a sedan slowed by the corner. She approached confidently as the passenger window rolled down.

"Looking for a good time, honey?" Jane said, leaning down to look inside the car. The man was not unattractive, with dark hair and nervous blue eyes. He nodded apprehensively. Jane smiled her warmest smile. She prided herself on her ability to reassure the customer, to make them feel comfortable and safe. The real trick of the trade, she knew, was to understand that safety and comfort were what they were really after.

"Do you ... you know, do the whole deal?" the man said, raising his eyebrows.

"No, honey. If that's what you're after you need to go down a few blocks. This is Unit Simulation only," she said, masking the disappointment.

"No, no," he said, shaking his head quickly, "that's what I want. Just the Simulation."

She smiled again. "Sure thing, darling. Just head around the corner. There's a parking structure there. Just park up and meet me at the alley around the back."

He nodded again and pulled away.

Jane walked to

the back of the building and waited. A few moments later, the man appeared from the structure, furtively glancing around to see if anyone was watching. He was a young professional type, dressed in a dark suit. Jane smiled a winsome smile.

"Don't worry, darling. Nobody comes around here. We're safe," she said. "My fee is one hundred for two hours." She paused with an expectant look for his assent. The man nodded. "Good. This way." Jane took the young man's hand and led him to a stairwell.

Upstairs they entered the studio. There was a couch with matching rug, a fireplace that gently crackled, books on a coffee table — all the little details of a family home from a bygone era, down to the children's toys that were casually strewn about.

"Just take off your coat and sit down," Jane said, indicating the couch. "I'll be right back."

The man complied and sat silently as Jane

left the room. In moments she returned, wearing pyjama bottoms and a cotton robe. Two children held her hands as she approached the man. They were Synthetics, but if you didn't look too closely it might

not be obvious. Jane knew that her customers didn't come here to examine things closely. They came to experience something that maybe their genes still remembered, and those types weren't interested in picking apart the details.

The man smiled hesitantly and the children smiled back. They hugged his neck and giggled. When they released him and went to play quietly on the rug with their toys, Jane sat on the couch beside the man. She placed a gentle arm around him and drifted into thought as they placidly watched the 'children' play, just as she always did.

There were advantages to the breeding programme, to be sure. Disease and deformity had nearly been eradicated, people were more intelligent, and the human race had finally begun to live in harmony with the planet, instead of at odds with it. But the price, Jane thought, was steep. Children, created and raised by the authorities, were perfect in every way, and perfectly sterile, just like Jane and her client. It had been generations since a true family had existed. Somewhere deep down, Jane knew, people felt that something was missing. The population was stable, meticulously regulated, but people could feel the absence of something lost.

Jane was proud. Even after Family Unit Simulation had been outlawed as a 'disturbance of moral judgment and psychological well-being', Jane had found a way to continue her trade. She knew what they really wanted, and she gave it to them, if only for two hours at a time.

The man looked at Jane with wet eyes and she pulled his head to her shoulder.

"There, there, darling. Everything is all right. You're home now." ■

Coy St Clair is a biologist and is presently putting the finishing touches on his first novel, *The Exile of William Bend*.

JACEY

